

---

## Time Series Models: Basics

### 12.1 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). A common simplifying assumption is that the data are equally spaced with a discrete-time observation index; however, this may only hold approximately. For example, daily log returns on a stock may only be available for weekdays, with additional gaps on holidays, and monthly values of the CPI are equally spaced by month, but unequally spaced by days. In either case, the consecutive observations are commonly regarded as equally spaced, for simplicity. 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 a 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 some form of distributional invariance over time, or *stationarity*, a property discussed next.

### 12.2 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.

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 stochastic behavior be constant in time. Often, it will suffice to assume 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

- $E(Y_t) = \mu$  (a finite constant) for all  $t$ ;
- $\text{Var}(Y_t) = \sigma^2$  (a positive finite constant) for all  $t$ ; and
- $\text{Cov}(Y_t, Y_s) = \gamma(|t - s|)$  for all  $t$  and  $s$  for some function  $\gamma(h)$ .

Thus, the mean and variance do not change with time and the covariance between two observations depends only on the *lag*, the time distance  $|t - s|$  between them, not the indices  $t$  or  $s$  directly. For example, if the process is weakly stationary, then the covariance between  $Y_2$  and  $Y_5$  is the same as the covariance 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. Weakly stationary is also sometimes referred to as *covariance stationary*. The term *stationary* will sometimes be used as a shorthand for strictly stationary.

The function  $\gamma$  is called the *autocovariance function* of the process. Note that  $\gamma(h) = \gamma(-h)$ . Why? Assuming weak stationarity, the correlation between  $Y_t$  and  $Y_{t+h}$  is denoted by  $\rho(h)$ . The function *rho* is called the *autocorrelation function*. Note that  $\gamma(0) = \sigma^2$  and that  $\gamma(h) = \sigma^2 \rho(h)$ . Also,  $\rho(h) = \gamma(h)/\sigma^2 = \gamma(h)/\gamma(0)$ .

As mentioned, many financial time series do not exhibit stationarity, but often the *changes* in them, perhaps after applying a log transformation, are approximately stationary. For this reason, stationary time series models have broad applicability and wide ranging applications. From the viewpoint of

---

<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.

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 Sect. 12.10.

A time series plot of a stationary series should show random 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.

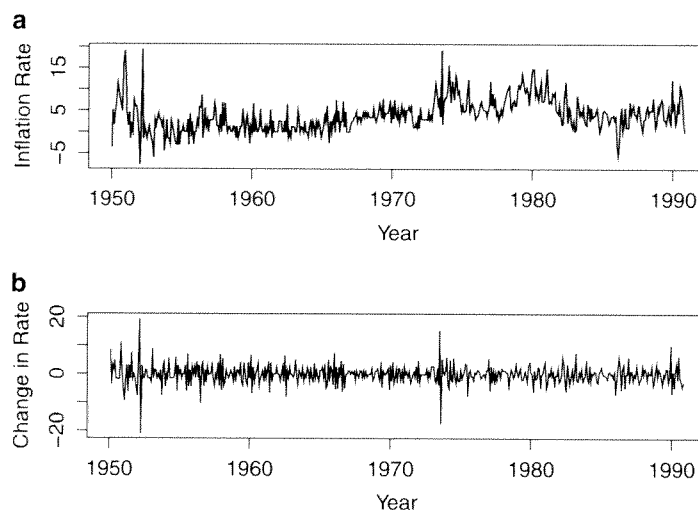
*Example 12.1. Inflation rates and changes in inflation rates—time series plots*

The one-month inflation rate (annual rate, in percent) is shown in Fig. 12.1a. The data come from the `Mishkin` data set in R's `Ecdat` package. The series may be wandering without reverting to a fixed mean, or it may be slowly reverting to a mean of approximately 4%, as would be expected with a stationary time series. In panel (b), the first differences, that is, the changes from one month to the next, are shown. In contrast to the original series, the differenced series certainly oscillate around a fixed mean that is 0%, or nearly so. The differenced series appears stationary, but whether or not the original series is stationary needs further investigation. We will return to this question later.  $\square$

*Example 12.2. Air passengers*

Figure 12.2 is a plot of monthly total international airline passengers for the years 1949 to 1960. The data come from the `AirPassengers` data set 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 with local peaks in summer and troughs in winter months, and third is the increase over time in the size of the seasonal oscillations.  $\square$



**Fig. 12.1.** Time series plots of (a) one-month inflation rate (annual rate, in percent) and (b) first differences (changes) in the one-month inflation rate. It is unclear if the series in (a) is stationary, but the differenced series in (b) seems suitable for modeling as stationary.

### 12.2.1 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  $\text{WN}(\mu, \sigma^2)$ ,” if

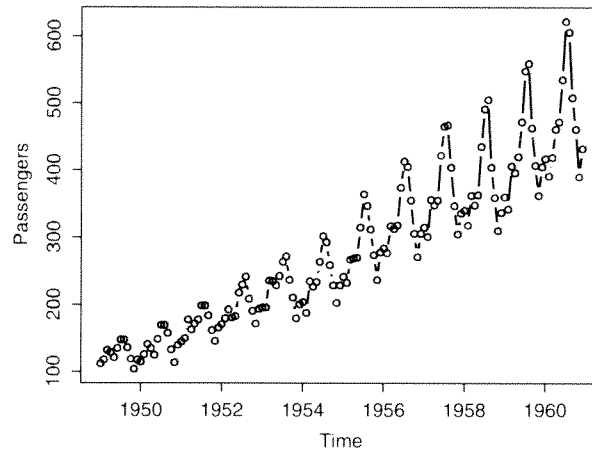
- $E(Y_t) = \mu$  (a finite constant) for all  $t$ ;
- $\text{Var}(Y_t) = \sigma^2$  (a positive finite constant) for all  $t$ ; and
- $\text{Cov}(Y_t, Y_s) = 0$  for all  $t \neq s$ .

If the mean is not specified, then it is assumed that  $\mu = 0$ . A weak white noise process is weakly stationary with

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

so that

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



**Fig. 12.2.** Time series plot of monthly totals of air passengers (in thousands).

If  $Y_1, Y_2, \dots$  is an i.i.d. process, then we call it an *i.i.d. white noise process* or simply *i.i.d. WN*( $\mu, \sigma^2$ ). Weak white noise is weakly stationary, while i.i.d. white noise is strictly stationary. An i.i.d. white noise process with  $\sigma^2$  finite 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.

### 12.2.2 Predicting White Noise

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

$$E(Y_{t+h} | Y_1, \dots, Y_t) = \mu \text{ for all } h \geq 1. \quad (12.1)$$

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  $Y_1, \dots, Y_t$  had not been observed. For weak white noise, (12.1) need not be true, but it is still true that the best linear predictor<sup>2</sup> of  $Y_{t+h}$  given  $Y_1, \dots, Y_t$  is  $\mu$ .

## 12.3 Estimating Parameters of a Stationary Process

Suppose we observe  $Y_1, \dots, Y_n$  from a weakly 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*

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (Y_{t+h} - \bar{Y})(Y_t - \bar{Y}) = n^{-1} \sum_{t=h+1}^n (Y_t - \bar{Y})(Y_{t-h} - \bar{Y}). \quad (12.2)$$

Equation (12.2) is an example of the usefulness of parsimony induced by the stationarity assumption. Because the covariance between  $Y_t$  and  $Y_{t+h}$  does not depend on  $t$ , *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 (12.2) 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

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

### 12.3.1 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.

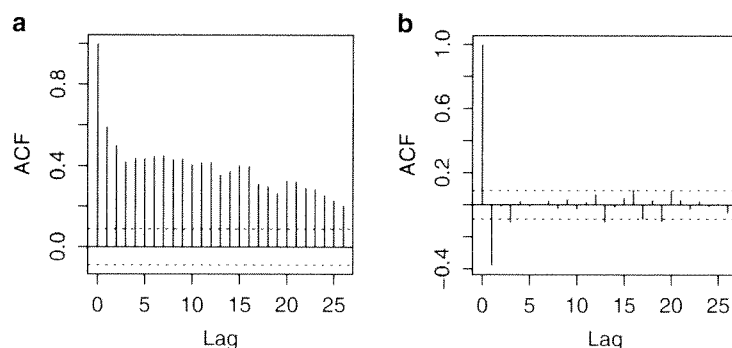
An alternative to using the bounds to test the autocorrelations one at a time is to use a simultaneous test. A *simultaneous test* is one that tests whether a group of null hypotheses are all true versus the alternative that at least one of them is false. The null hypothesis of the Ljung–Box test is  $H_0 : \rho(1) = \rho(2) = \dots = \rho(K) = 0$  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.

<sup>2</sup> Best linear prediction is discussed in Sect. 11.9.1.

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 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 12.3. Inflation rates and changes in the inflation rate—sample ACF plots and the Ljung-Box test*



**Fig. 12.3.** Sample ACF plots of the one-month inflation rate (a) and changes in the inflation rate (b).

We return to the inflation rate data used in Example 12.1. Figure 12.3 contains plots of (a) the sample ACF of the one-month inflation rate and (b) the sample ACF of changes in the inflation rate.

```
1 data(Mishkin, package = "Ecdat")
2 y = as.vector(Mishkin[,1])
3 par(mfrow=c(1,2))
4 acf(y)
5 acf(diff(y))
```

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 Sect. 13.5. 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 Fig. 12.1, specifically that the differenced series is stationary and the original series might not be. In Sect. 12.10 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 change 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.  $K$  is called `lag` when `Box.test()` is called and `df` in the output, and we specify `type="Ljung-Box"` for the Ljung–Box test.

```
6 Box.test(diff(y), lag = 10, type = "Ljung-Box")
```

The Ljung–Box test statistic with  $K = 10$  is 79.92, which has an extremely small  $p$ -value,  $5.217\text{e}-13$ , so the null hypothesis of white noise is strongly rejected. Other choices of  $K$  give similar results.  $\square$

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.

## 12.4 AR(1) Processes

Time series models with correlation can be constructed from 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 weak  $\text{WN}(0, \sigma_\epsilon^2)$ . We say that  $Y_1, Y_2, \dots$  is an *AR(1) process* if for some constant parameters  $\mu$  and  $\phi$ ,

$$Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t \quad (12.3)$$

for all  $t$ . The parameter  $\mu$  is the mean of the process, hence  $(Y_t - \mu)$  has mean zero for all  $t$ . We may interpret 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 weak  $\text{WN}(\mu, \sigma_\epsilon^2)$ . 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  $Y_1, Y_2, \dots$  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 (12.3) 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 (12.3) can be rewritten as

$$Y_t = (1 - \phi)\mu + \phi Y_{t-1} + \epsilon_t. \quad (12.4)$$

Recall the linear regression model  $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$  from your statistics courses or see Chap. 9 for an introduction to regression analysis. Equation (12.4) is just a linear regression model with intercept  $\beta_0 = (1 - \phi)\mu$  and slope  $\beta_1 = \phi$ . The term *autoregression* refers to the regression of the process on its own past values.

If  $|\phi| < 1$ , then repeated substitution of (12.3) shows that

$$Y_t = \mu + \epsilon_t + \phi \epsilon_{t-1} + \phi^2 \epsilon_{t-2} + \dots = \mu + \sum_{h=0}^{\infty} \phi^h \epsilon_{t-h}, \quad (12.5)$$

assuming that time index  $t$  of  $Y_t$  and  $\epsilon_t$  can be extended to negative values so that the white noise process is  $\{\epsilon_t\}_{t=-\infty}^{+\infty}$  and (12.3) is true for all integers  $t$ . Equation (12.5) 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 (12.5) could be truncated at a finite number of terms, with no practical need to assume that the processes existed in the infinite past.

### 12.4.1 Properties of a Stationary AR(1) Process

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

$$E(Y_t) = \mu \quad \forall t, \quad (12.6)$$

$$\text{Var}(Y_t) = \gamma(0) = \sigma_Y^2 = \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \forall t, \quad (12.7)$$

$$\text{Cov}(Y_t, Y_{t+h}) = \gamma(h) = \phi^{|h|} \frac{\sigma_\epsilon^2}{1 - \phi^2} \quad \forall t \text{ and } \forall h, \text{ and} \quad (12.8)$$

$$\text{Corr}(Y_t, Y_{t+h}) = \rho(h) = \phi^{|h|} \quad \forall t \text{ and } \forall h. \quad (12.9)$$

It is important to remember that formulas (12.6) to (12.9) 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 (12.5). For example, using (7.11) in Sect. 7.3.2,

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{h=0}^{\infty} \phi^h \epsilon_{t-h}\right) = \sigma_{\epsilon}^2 \sum_{h=0}^{\infty} \phi^{2h} = \frac{\sigma_{\epsilon}^2}{1 - \phi^2}, \quad (12.10)$$

which proves (12.7). In (12.10) the formula for summation of a geometric series was used. This formula is

$$\sum_{i=0}^{\infty} r^i = \frac{1}{1-r} \quad \text{if } |r| < 1. \quad (12.11)$$

Also, for  $h > 0$ ,

$$\text{Cov}\left(\sum_{i=0}^{\infty} \epsilon_{t-i} \phi^i, \sum_{j=0}^{\infty} \epsilon_{t+h-j} \phi^j\right) = \phi^{|h|} \frac{\sigma_{\epsilon}^2}{1 - \phi^2}, \quad (12.12)$$

thus verifying (12.8). Then (12.9) follows by dividing (12.8) by (12.7).

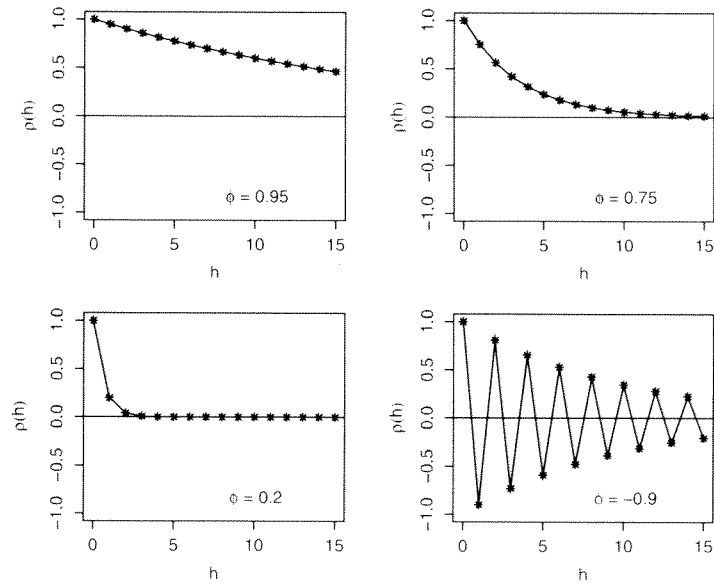
Be sure to distinguish between  $\sigma_{\epsilon}^2$ , which is the variance of the white noise process  $\epsilon_1, \epsilon_2, \dots$ , and  $\gamma(0)$ , which is the variance,  $\sigma_Y^2$ , of the stationary AR(1) process  $Y_1, Y_2, \dots$ . We can see from (12.7) that  $\gamma(0)$  is larger than  $\sigma_{\epsilon}^2$  unless  $\phi = 0$ , in which case  $Y_t = \mu + \epsilon_t$ , such that  $Y_t$  and  $\epsilon_t$  have the same variance.

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 Fig. 12.4. 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, or 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.

### 12.4.2 Convergence to the Stationary Distribution

Suppose that  $Y_0$  is an arbitrary starting value not chosen from the stationary distribution and that (12.3) holds for  $t = 1, 2, \dots$ . Then the process is not stationary, but converges to the stationary distribution satisfying (12.6) to (12.9) as  $t \rightarrow \infty$ .<sup>3</sup> For example, since  $Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$ , we have

<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  $t > 0$ .



**Fig. 12.4.** Autocorrelation functions of AR(1) processes with  $\phi$  equal to 0.95, 0.75, 0.2, and  $-0.9$ .

$E(Y_1) - \mu = \phi\{E(Y_0) - \mu\}$ , and  $E(Y_2) - \mu = \phi^2\{E(Y_0) - \mu\}$ , and so forth, so that

$$E(Y_t) = \mu + \phi^t\{E(Y_0) - \mu\} \text{ for all } t > 0. \quad (12.13)$$

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 (12.13),  $E(Y_{10})$  is very close to  $\mu$  unless  $E(Y_0)$  was extremely far from  $\mu$ .

### 12.4.3 Nonstationary AR(1) Processes

If  $|\phi| \geq 1$ , then the AR(1) process is nonstationary, and the mean, variance, covariances and and correlations 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 Chap. 2.

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

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

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.

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

When  $|\phi| > 1$ , an AR(1) process has explosive behavior. This can be seen in Fig. 12.5. 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  $\phi = 1$ .

## 12.5 Estimation of AR(1) Processes

R has the function `arima()` for fitting AR and other time series models. The function `arima()` and similar functions in other software packages have two primary estimation methods, conditional least-squares and maximum likelihood. The two methods are explained in Sect. 12.5.2. They are 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 used as the starting value for computing the MLE by iterative non-linear optimization.

### 12.5.1 Residuals and Model Checking

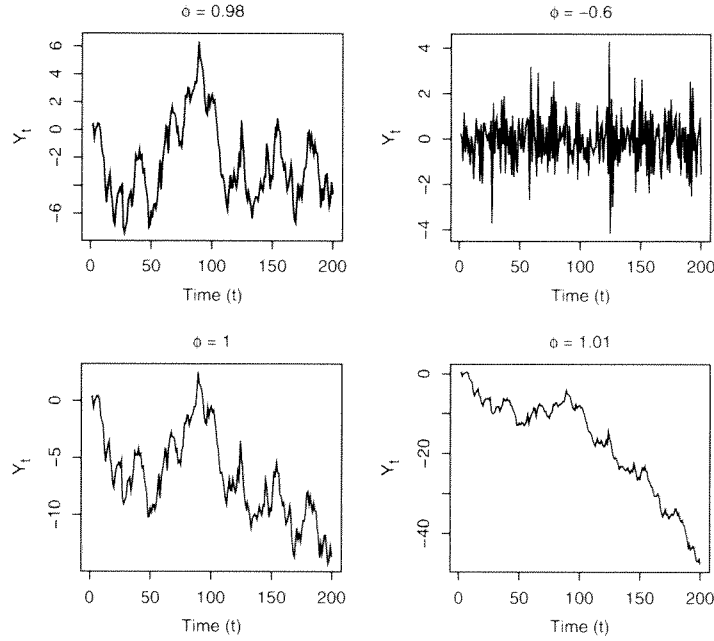
Once  $\mu$  and  $\phi$  have been estimated, one can estimate the white noise process  $\epsilon_1, \dots, \epsilon_n$ . Rearranging Eq. (12.3), we have

$$\epsilon_t = (Y_t - \mu) - \phi(Y_{t-1} - \mu). \quad (12.14)$$

In analogy with (12.14), the residuals,  $\hat{\epsilon}_2, \hat{\epsilon}_3, \dots, \hat{\epsilon}_n$ , are defined as

$$\hat{\epsilon}_t = (Y_t - \hat{\mu}) - \hat{\phi}(Y_{t-1} - \hat{\mu}), \quad t \geq 2, \quad (12.15)$$

and estimate  $\epsilon_2, \dots, \epsilon_n$ . The first noise,  $\epsilon_1$ , cannot be estimated directly since it is assumed that the observations start at  $Y_1$  so that  $Y_0$  is not available.



**Fig. 12.5.** 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.

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

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

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

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 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 autoregressive coefficient parameters that were estimated, so for an AR(1) model `fitdf` should be 1.

*Example 12.4. Daily log returns for BMW stock—ACF plots and AR fit*

The daily log returns for BMW stock between January 1973 and July 1996 from the `bmw` data set in R's `evir` package are shown in Fig. 12.6a. Their sample ACF and quantiles are shown in Fig. 12.6b and c, respectively. The estimated 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 lag autocorrelations are 0 was performed using R's `Box.test()` function.

```
7 data(bmw, package = "evir")
8 Box.test(bmw, lag = 5, type = "Ljung-Box")
```

The parameter `lag`, which 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

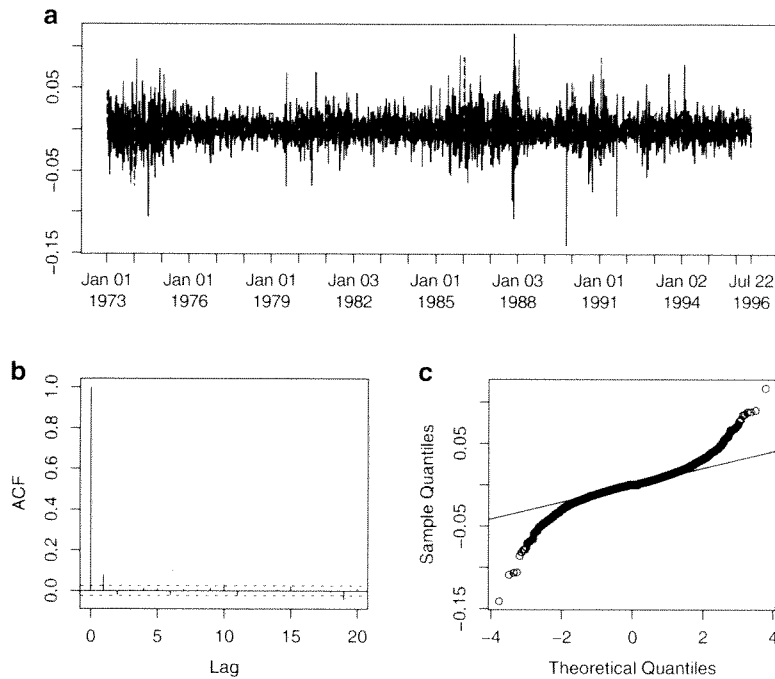
```
data:  bmw
X-squared = 44.987, df = 5, p-value = 1.460e-08
```

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

Next, an AR(1) model was fit using the `arima()` command in R.

```
9 fitAR1 = arima(bmw, order = c(1,0,0))
10 print(fitAR1)
```

The `order` parameter will be explained later, but for an AR(1) process it should be `c(1,0,0)`. A summary of the output is below.



**Fig. 12.6.** (a) Daily log returns for BMW stock from January 1973 until July 1996, and their (b) sample ACF and (c) sample quantiles relative to the normal distribution.

Call:

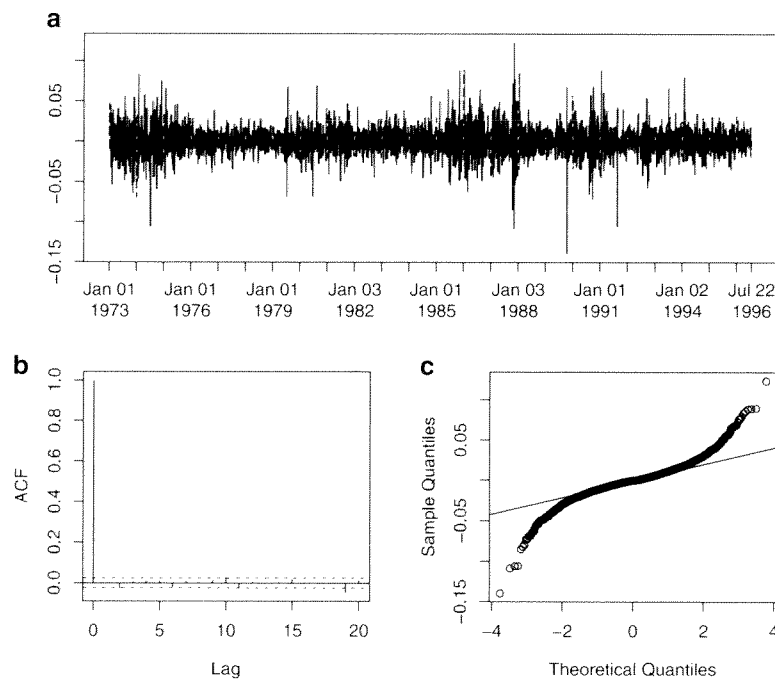
```
arima(x = bmw, order = c(1, 0, 0))
```

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  $\hat{\sigma}^2 = 0.00022$ . Although  $\hat{\phi}$  is small, it is statistically significant since it is 6.4 times its standard error 0.013, 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 non-zero 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.



**Fig. 12.7.** A (a) time series plot, (b) sample ACF and (c) normal quantile plot of residuals from an  $AR(1)$  fit to the daily log returns for BMW stock.

The sample ACF of the residuals is plotted in Fig. 12.7b. None of the autocorrelations at low lags are 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 `lag` equal to 5 and `fitdf=1`.

```
11 Box.test(residuals(fitAR1), lag = 5, type = "Ljung-Box", fitdf = 1)
```

Box-Ljung test

```
data: residuals(fitAR1)
X-squared = 6.8669, df = 4, 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 `lag` equal to 10, 15, and 20 and the  $p$ -values were 0.041, 0.045, and 0.004, 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 time series plot and normal quantile plot of the AR(1) residuals in Fig. 12.7a and c show volatility clustering and heavy tails. These are common features of economic data and will be modeled in subsequent chapters.  $\square$

*Example 12.5. Inflation rate—AR(1) fit and checking residuals*

This example uses the inflation rate time series used earlier in Example 12.1. 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 Fig. 12.8 and there is considerable residual autocorrelation, which indicates that the AR(1) model is not adequate. A Ljung–Box test confirms this result.

```

12 data(Mishkin, package = "Ecdat")
13 y = as.vector(Mishkin[,1])
14 fit = arima(y, order = c(1,0,0))
15 Box.test(fit$resid, type = "Ljung", lag = 24, fitdf = 1)

```

Box-Ljung test

```

data: fit$resid
X-squared = 138.5776, df = 23, p-value < 2.2e-16

```

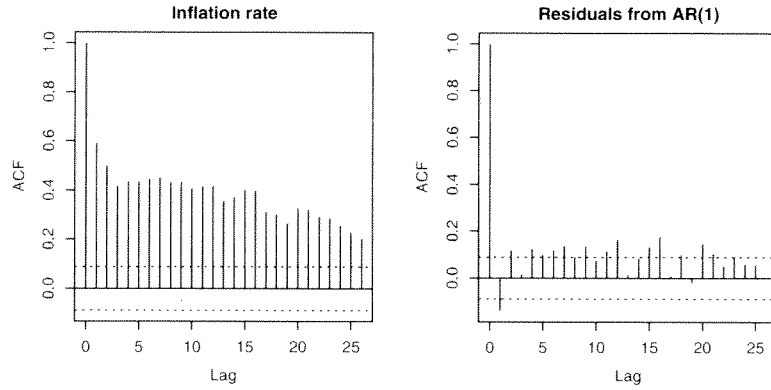
One might try fitting an AR(1) to the changes in the inflation rate, since this series is clearly stationary. However, the AR(1) model also does not fit the changes in the inflation rate. We will return to this example when we have a larger collection of models in our statistics toolbox.  $\square$

### 12.5.2 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 joint density for  $Y_1, \dots, Y_n$ , we use (A.41) and the fact that

$$f_{Y_t|Y_1, \dots, Y_{t-1}}(y_t|y_1, \dots, y_{t-1}) = f_{Y_t|Y_{t-1}}(y_t|y_{t-1}) \quad (12.16)$$



**Fig. 12.8.** Sample ACF for the inflation rate time series and residual series from an AR(1) fit.

for  $t = 2, 3, \dots, n$ . A stochastic process with property (12.16) is called a *Markov process*. By (A.41) and (12.16), we have

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{Y_1}(y_1) \prod_{t=2}^n f_{Y_t|Y_{t-1}}(y_t|y_{t-1}). \quad (12.17)$$

If we assume the errors are from a Gaussian white noise process, then by (12.6) and (12.7), we know that  $Y_1$  is  $N\{\mu, \sigma_\epsilon^2/(1 - \phi^2)\}$  for a stationary AR(1) process. Given  $Y_{t-1}$ , the only random component of  $Y_t$  is  $\epsilon_t$ , so that  $Y_t$  given  $Y_{t-1}$  is  $N\{\mu + \phi(Y_{t-1} - \mu), \sigma_\epsilon^2\}$ . It then follows that the joint density for  $Y_1, \dots, Y_n$  is

$$\left(\frac{1}{\sqrt{2\pi}\sigma_\epsilon}\right)^n \sqrt{1 - \phi^2} \exp\left\{-\frac{(Y_1 - \mu)^2}{2\sigma_\epsilon^2/(1 - \phi^2)}\right\} \prod_{i=2}^n \exp\left(-\frac{\left[Y_i - \{\mu + \phi(Y_{i-1} - \mu)\}\right]^2}{2\sigma_\epsilon^2}\right). \quad (12.18)$$

The maximum likelihood estimator maximizes the logarithm of (12.18) over  $(\mu, \phi, \sigma_\epsilon)$ . A somewhat simpler estimator deletes the marginal density of  $Y_1$  from the likelihood and maximizes the logarithm of

$$\left(\frac{1}{\sqrt{2\pi}\sigma_\epsilon}\right)^{n-1} \prod_{t=2}^n \exp\left(-\frac{\left[Y_t - \{\mu + \phi(Y_{t-1} - \mu)\}\right]^2}{2\sigma_\epsilon^2}\right). \quad (12.19)$$

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

$$\sum_{t=2}^n \left[ Y_t - \{ \mu + \phi(Y_{t-1} - \mu) \} \right]^2. \quad (12.20)$$

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, along with approximate standard errors. The default option is used in the examples in this book.

## 12.6 AR( $p$ ) Models

We have seen that the ACF of an AR(1) process decays geometrically to zero if  $|\phi| < 1$  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 models, but one that is still 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 is 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  $\epsilon_1, \epsilon_2, \dots$  is weak 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 also be expressed 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 the “References” section. However, software is available for computing and plotting the ACF of any AR processes, as well as for the MA and ARMA processes to be introduced soon. Figure 12.9 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

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

<sup>5</sup> See Chap. 9 for an introduction to multiple regression.

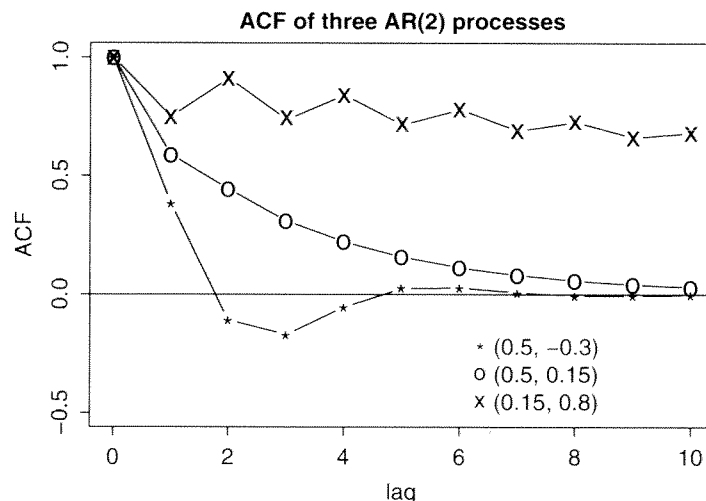


Fig. 12.9. ACF of three AR(2) processes; the legend gives the values of  $\phi_1$  and  $\phi_2$ .

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.

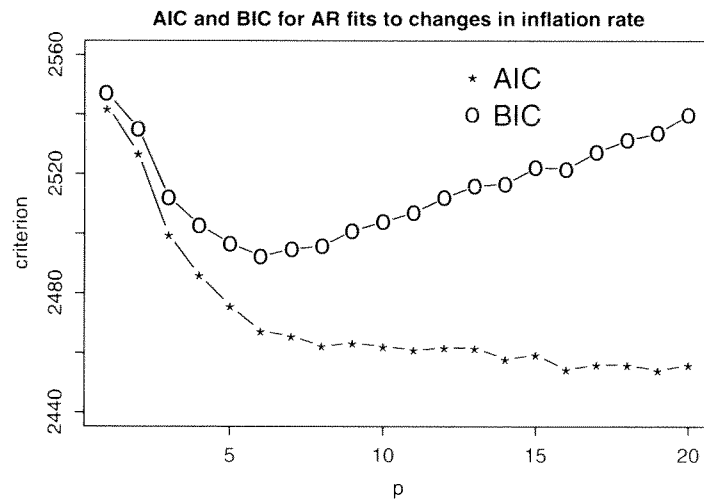
*Example 12.6. Changes in the inflation rate—AR( $p$ ) models*

Figure 12.10 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 a MA or ARMA model, which are the 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.

```
16 library(forecast)
17 auto.arima(diff(y), max.p = 20, max.q = 0, d = 0, ic = "aic")
```

```
Series: diff(y)
ARIMA(8,0,0) with zero mean
```



**Fig. 12.10.** Fitting AR( $p$ ) models to changes in the one-month inflation rate: AIC and BIC plotted against  $p$ .

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6
	-0.6274	-0.4977	-0.5158	-0.4155	-0.3443	-0.2560
s.e.	0.0456	0.0536	0.0576	0.0606	0.0610	0.0581

	ar7	ar8
	-0.1557	-0.1051
s.e.	0.0543	0.0459

sigma<sup>2</sup> estimated as 8.539: log likelihood=-1221.2  
 AIC=2460.4 AICc=2460.78 BIC=2498.15

The first local minimum of BIC is at  $p = 6$ .

```
18 auto.arima(diff(y), max.p = 20, max.q = 0, d = 0, ic = "bic")
```

Series: diff(y)  
 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<sup>2</sup> estimated as 8.699: log likelihood=-1225.67  
 AIC=2465.33 AICc=2465.56 BIC=2494.69

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

*Example 12.7. 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( $p$ ) model was fit to the inflation rates with  $p$  selected via an information criterion by `auto.arima()`. The BIC method chose  $p = 2$  and AIC selected  $p = 7$ . The results for  $p = 7$  are below.

```
Series: y
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         3.987
s.e.    0.045  0.048   0.048  0.048  0.049  0.048  0.046         0.784

sigma^2 estimated as 8.47:  log likelihood=-1221.8
AIC=2461.6   AICc=2461.9   BIC=2499.3
```

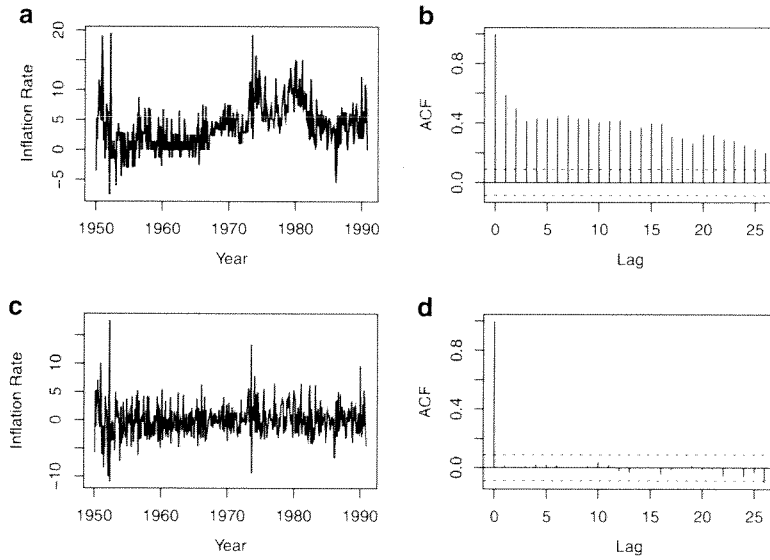
The inflation rate and its residual series from an AR(7) fit and their sample ACFs are shown in Fig. 12.11. □

## 12.7 Moving Average (MA) Processes

As we saw in Example 12.6, 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.

### 12.7.1 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, Fig. 12.3b 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 12.6, 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  $\{\epsilon_t\}$ .



**Fig. 12.11.** (a) The inflation rate series and (b) its sample ACF; (c) the residuals series from an  $AR(7)$  fit to the inflation rates (d) ACF of residuals.

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

$$Y_t - \mu = \epsilon_t + \theta\epsilon_{t-1}, \quad (12.21)$$

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

One can show that

$$\begin{aligned} E(Y_t) &= \mu, \\ \text{Var}(Y_t) &= \sigma_\epsilon^2(1 + \theta^2), \\ \gamma(1) &= \theta\sigma_\epsilon^2, \\ \gamma(h) &= 0 \text{ if } |h| > 1, \\ \rho(1) &= \frac{\theta}{1 + \theta^2}, \end{aligned} \quad (12.22)$$

$$\rho(h) = 0 \text{ if } |h| > 1. \quad (12.23)$$

Notice the implication of (12.22) and (12.23)—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.

<sup>6</sup> Some textbooks and some software write MA models with the signs reversed so that model (12.21) 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 Eqs. (12.24) and (12.25).

### 12.7.2 General MA Processes

The  $\text{MA}(q)$  process is

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}. \quad (12.24)$$

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  $\text{AR}(p)$  models where the “constant” in the model is not the same as the mean, in an  $\text{MA}(q)$  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 Eq. (12.24), 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.  $\text{MA}(q)$  models can be fit easily using, for example, the `arima()` function in R.

#### Example 12.8. Changes in the inflation rate—MA models

$\text{MA}(q)$  models were fit to the changes in the inflation rate. Figure 12.12 shows plots of AIC and BIC versus  $q$ . BIC suggests that an  $\text{MA}(2)$  model is adequate, while AIC suggests an  $\text{MA}(3)$  model. We fit the  $\text{MA}(3)$  model. The Ljung–Box test was applied to the residuals with `fitdf = 3` and `lag` equal to 5, 10, and 15 and gave  $p$ -values of 0.65, 0.76, and 0.32, respectively. The  $\text{MA}(2)$  also provided an adequate fit with the  $p$ -values from the Ljung–Box test all above 0.08. The output for the  $\text{MA}(3)$  model is below.

```

19 fitMA3 = arima(diff(y), order = c(0,0,3))
20 fitMA3

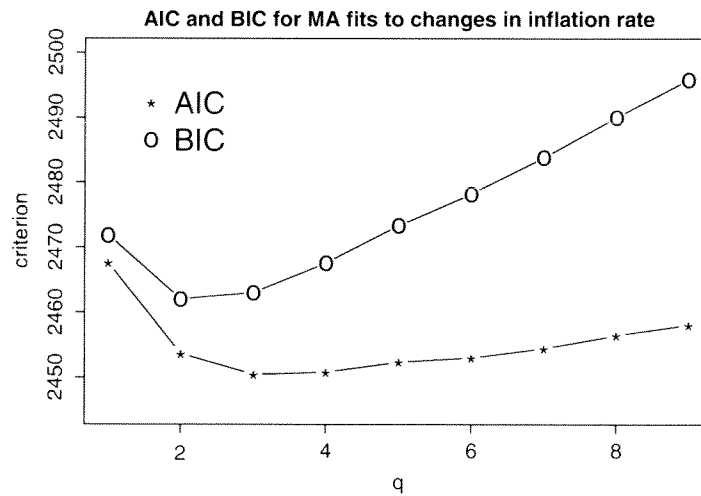
Series: diff(y)
ARIMA(0,0,3) with non-zero mean

Coefficients:
          ma1      ma2      ma3  intercept
      -0.633  -0.103  -0.108         0.000
s.e.    0.046   0.051   0.047         0.021

sigma^2 estimated as 8.5:  log likelihood=-1220.3
AIC=2450.5   AICc=2450.7   BIC=2471.5
```

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.  $\square$





**Fig. 12.12.** Fitting  $MA(q)$  models to changes in the one-month inflation rate; AIC and BIC plotted against  $q$ .

## 12.8 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.

### 12.8.1 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,

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

Thus,  $B$  backs up time one unit while  $B^h$  does this repeatedly so that time is backed up  $h$  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*.

### 12.8.2 The ARMA Model

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

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}, \quad (12.25)$$

which shows how  $Y_t$  depends on lagged values of itself and lagged values of the white noise process. Equation (12.25) can be written more succinctly with the backwards operator as

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(Y_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t. \quad (12.26)$$

A white noise process is  $ARMA(0,0)$  since if  $p = q = 0$ , then (12.26) reduces to

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

### 12.8.3 ARMA(1,1) Processes

The  $ARMA(1,1)$  model is commonly used in practice and is simple enough to study theoretically. In this 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

$$Y_t = \phi Y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \quad (12.27)$$

by  $\epsilon_t$  and taking expectations, one has

$$\text{Cov}(Y_t, \epsilon_t) = E(Y_t \epsilon_t) = \sigma_\epsilon^2, \quad (12.28)$$

since  $\epsilon_t$  is independent of  $\epsilon_{t-1}$  and  $Y_{t-1}$ . From (12.27) and (12.28),

$$\gamma(0) = \phi^2 \gamma(0) + (1 + \theta^2) \sigma_\epsilon^2 + 2\phi\theta \sigma_\epsilon^2, \quad (12.29)$$

and then solving (12.29) for  $\gamma(0)$  gives us the formula

$$\gamma(0) = \frac{(1 + \theta^2 + 2\phi\theta) \sigma_\epsilon^2}{1 - \phi^2}. \quad (12.30)$$

By similar calculations, multiplying (12.27) by  $Y_{t-1}$  and taking expectations yields a formula for  $\gamma(1)$ . Dividing this formula by the right-hand side of (12.29) gives us

$$\rho(1) = \frac{(1 + \phi\theta)(\phi + \theta)}{1 + \theta^2 + 2\phi\theta}. \quad (12.31)$$

For  $h \geq 2$ , multiplying (12.27) by  $Y_{t-h}$  and taking expectations results in the formula

$$\rho(h) = \phi \rho(h-1), \quad h \geq 2. \quad (12.32)$$

By (12.32), 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  $\phi$ .

### 12.8.4 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 Sect. 12.5. 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 12.9. Changes in risk-free returns—ARMA models*

This example uses the monthly changes in the risk-free returns shown in Fig. 4.3. In Table 12.1, 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.

Sample ACF, normal, and time series plots of the residuals from the ARMA(1,1) model are shown in Fig. 12.13. 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 residual time series plot shows volatility clustering. These problems will be addressed in later chapters.  $\square$

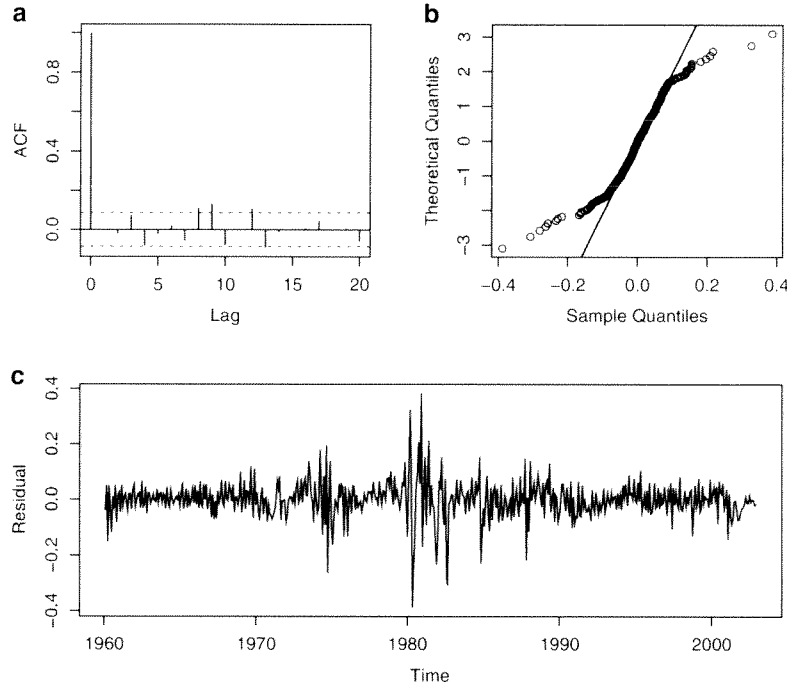
**Table 12.1.** *AIC and BIC for ARMA models fit to the monthly changes in the risk-free interest returns. The minimum values of both criteria are shown in boldface. To improve the appearance of the table, 1290 was added to all AIC and BIC values.*

$p$	$q$	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

### 12.8.5 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. 12.13.** Residual sample ACF, normal quantile, and time series 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:

$$\Delta^k Y_t = (1 - B)^k Y_t = \sum_{\ell=0}^k \binom{k}{\ell} (-1)^\ell Y_{t-\ell}. \quad (12.33)$$

## 12.9 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 this 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, and white noise is an  $ARIMA(0, 0, 0)$  model.

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

$$w_t = w_{t_0} + Y_{t_0+1} + \cdots + Y_t. \quad (12.34)$$

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

$$\Delta w_t = Y_t, \quad (12.35)$$

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

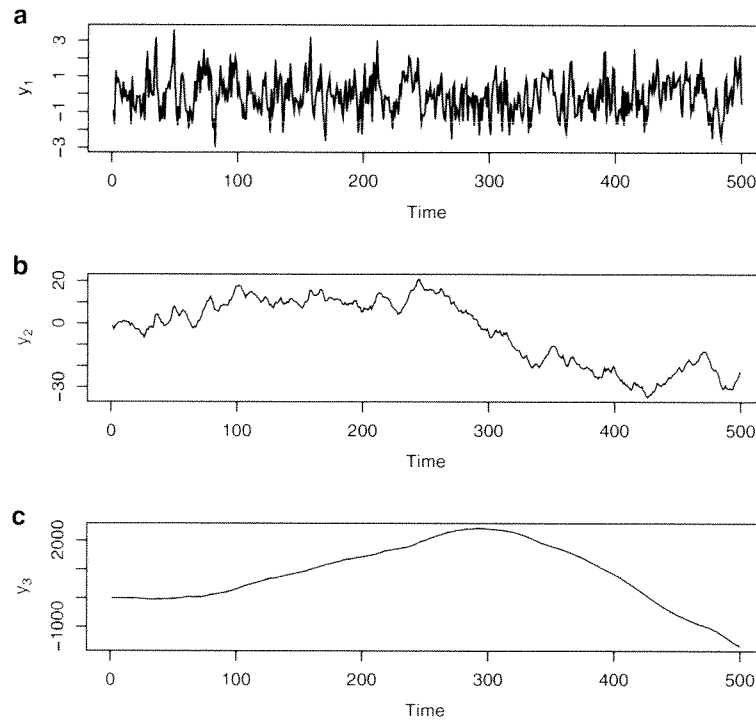
We will say that a process is  $I(d)$  if it is stationary after being differenced  $d$  times. For example, a stationary process is  $I(0)$ . An  $ARIMA(p, d, q)$  process is  $I(d)$ . An  $I(d)$  process is said to be “integrated to order  $d$ .”

Figure 12.14 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 `arma.sim()`. This process was integrated twice with R’s `cumsum()` function.

```
21 set.seed(4631)
22 y1 = arma.sim(n = 500, list(ar = c(0.4)))
23 y2 = cumsum(y1)
24 y3 = cumsum(y2)
```

*Example 12.10. Fitting an ARIMA model to CPI data*

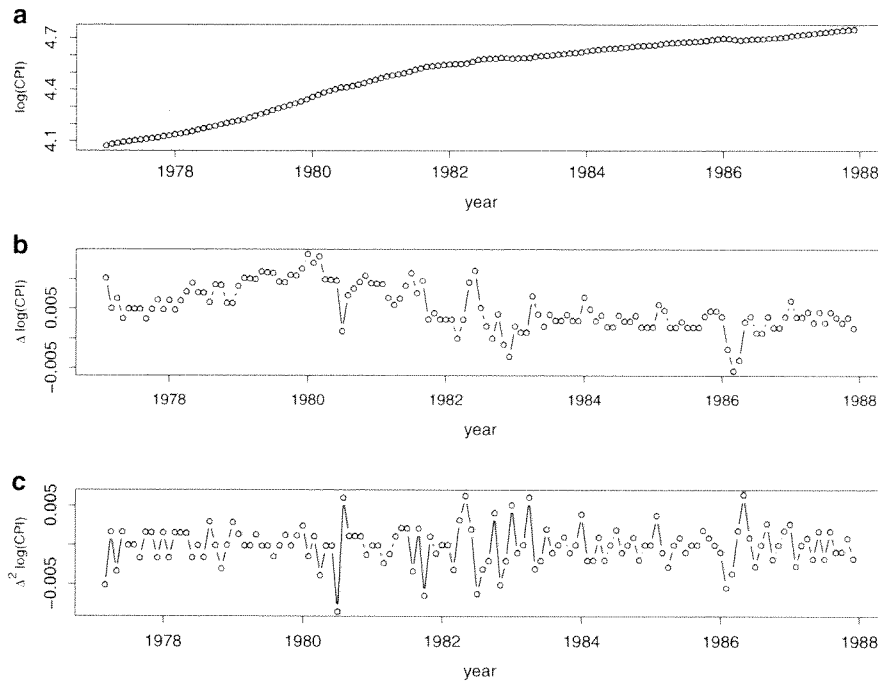
<sup>7</sup> An analog is, of course, differentiation and integration in calculus, which are inverses of each other.



**Fig. 12.14.** The (a) top plot is of an  $AR(1)$  process with  $\mu = 0$  and  $\phi = 0.4$ . The (b) middle and (c) 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.

This example uses the `CPI.dat.csv` data set. 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 12.15 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 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 Fig. 12.16a,b, and c provide additional evidence that the  $\log(\text{CPI})$  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 ACF of the residuals from this fit is shown in Fig. 12.16d. 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 `lag`, for example, with a  $p$ -value of 0.08 at `lag = 20`, with `fitddf = 2`. □



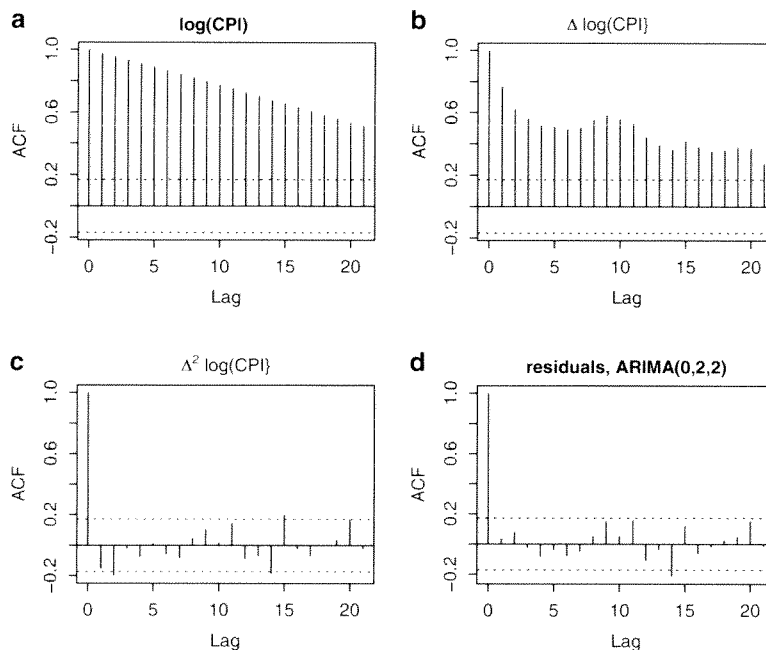
**Fig. 12.15.** (a)  $\log(\text{CPI})$ , (b) first differences of  $\log(\text{CPI})$ , and (c) second differences of  $\log(\text{CPI})$ .

*Example 12.11. Fitting an ARIMA model to industrial production (IP) data*

This example uses the `IP.dat` data set. The variable, `IP`, is a seasonally adjusted U.S. industrial production index. Figure 12.17 panels (a) and (b) show time series plots of  $\log(\text{IP})$  and  $\Delta \log(\text{IP})$  and panel (c) has the sample ACF of  $\Delta \log(\text{IP})$ . The  $\log(\text{IP})$  series appears to be  $I(1)$ , implying that we should fit an ARMA model to  $\Delta \log(\text{IP})$ .  $\text{AR}(1)$ ,  $\text{AR}(2)$ , and  $\text{ARMA}(1,1)$  each fit  $\Delta \log(\text{IP})$  reasonably well and the  $\text{ARMA}(1,0)$  model is selected using the BIC criterion with R's `auto.arima()` function. The ACF of the residuals in Fig. 12.17d indicates a satisfactory fit to the  $\text{ARMA}(1,0)$  model since it shows virtually no short-term autocorrelation. In summary,  $\log(\text{IP})$  is well fit by an  $\text{ARIMA}(1,1,0)$  model.  $\square$

### 12.9.1 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.



**Fig. 12.16.** Sample 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  $\text{ARIMA}(0,2,2)$  model fit to  $\log(\text{CPI})$ .

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 Sect. 12.11 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  $E(\Delta^d Y_t)$  has an  $(m - d)$ th-degree trend for  $d \leq m$ . For  $d > m$ ,  $E(\Delta^d Y_t) = 0$ .

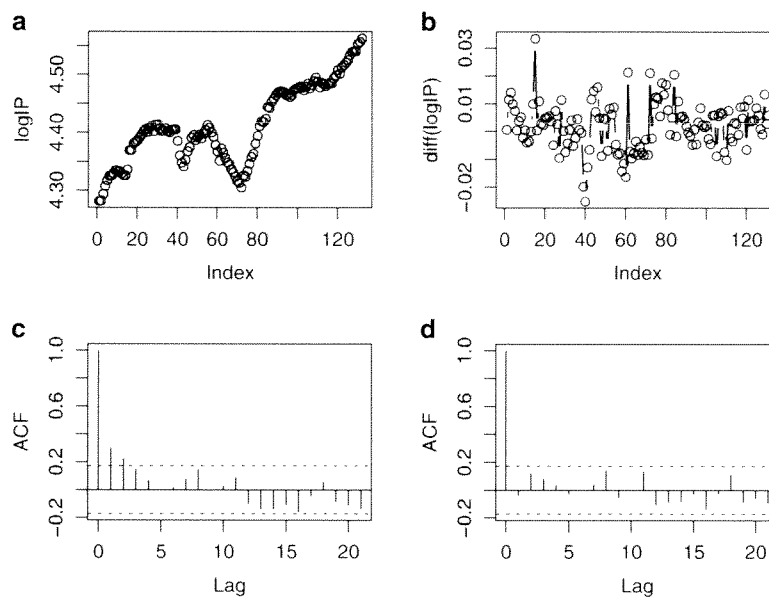
## 12.10 Unit Root Tests

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  $\text{ARMA}(p, q)$  process can be written as

$$(Y_t - \mu) = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}. \quad (12.36)$$





**Fig. 12.17.** Time series plots of (a)  $\log(IP)$  and (b)  $\Delta \log(IP)$ , and sample ACF plots of (c)  $\Delta \log(IP)$  and (d) residual from  $ARMA(1,0)$  fit to  $\Delta \log(IP)$ .

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

$$1 - \phi_1 x - \cdots - \phi_p x^p \quad (12.37)$$

have absolute values greater than one. (See Appendix A.21 for information about complex roots of polynomials and the absolute value of a complex number.) For example, when  $p = 1$ , then (12.37) is

$$1 - \phi x$$

and has one root,  $1/\phi$ . 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 12.12. Inflation rates

Recall from Examples 12.1 and 12.3 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  $\hat{\phi}_1 = 1.229$  and  $\hat{\phi}_2 = -0.233$ . The roots of

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

can be found easily using R's `polyroot()` function and are 1.0053 and 4.2694.

```
25 polyroot(c(1, -1.229, +0.233))
```

Both roots have absolute values greater than 1, indicating possible stationarity; however, 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 12.13. 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 `adf.test()`, `pp.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()`.

```
26 library(tseries)
```

```
27 adf.test(y)
```

```
Augmented Dickey-Fuller Test
```

```
data: y
```

```
Dickey-Fuller = -3.8651, Lag order = 7, p-value = 0.01576
```

```
alternative hypothesis: stationary
```

```
28 pp.test(y)
```

```
Phillips-Perron Unit Root Test
```

```
data: y
```

```
Dickey-Fuller Z(alpha) = -248.75, Truncation lag parameter = 5,
```

```
p-value = 0.01
```

```
alternative hypothesis: stationary
```

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.

```
29 kpss.test(y)

KPSS Test for Level Stationarity
data: y
KPSS Level = 2.51, Truncation lag parameter = 5, p-value = 0.01
```

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 Sect. 13.5.  $\square$

### 12.10.1 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 Sect. 12.14 for more information. The Dickey–Fuller test is based on the AR(1) model

$$Y_t = \phi Y_{t-1} + \epsilon_t. \quad (12.38)$$

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  $\phi > -1$ . The AR(1) model (12.38) is equivalent to  $\Delta Y_t = (\phi - 1)Y_{t-1} + \epsilon_t$ , or

$$\Delta Y_t = \pi Y_{t-1} + \epsilon_t, \quad (12.39)$$

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 (12.39) by adding a time trend and lagged values of  $\Delta Y_t$ . Typically, the time trend is linear so that the expanded model is

$$\Delta Y_t = \beta_0 + \beta_1 t + \pi Y_{t-1} + \sum_{j=1}^p \gamma_j \Delta Y_{t-j} + \epsilon_t. \quad (12.40)$$

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(y)-1)^(1/3))`, where  $y$  is the input series ( $Y_t$  in our notation).

### 12.11 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 an 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 12.14. 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.

```
30 auto.arima(y, max.p = 5, max.q = 5, ic = "bic", trace = FALSE)
```

```
Series: y
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=-1221.6
AIC=2449.2  AICc=2449.3  BIC=2461.8
```

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

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

### 12.12 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

$$Y_{n+1} = \mu + \phi(Y_n - \mu) + \epsilon_{n+1}. \quad (12.41)$$

Since  $\epsilon_{n+1}$  is independent of the past and present, by Result 11.1 in Sect. 11.9.2 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 (12.41). 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

$$\hat{Y}_{n+2} = \hat{\mu} + \hat{\phi}(\hat{Y}_{n+1} - \hat{\mu}) = \hat{\mu} + \hat{\phi}\{\hat{\phi}(Y_n - \hat{\mu})\}, \quad (12.42)$$

and so forth. Notice that in (12.42) 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:

$$\hat{Y}_{n+k} = \hat{\mu} + \hat{\phi}^k(Y_n - \hat{\mu}). \quad (12.43)$$

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

Formula (12.43) 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

$$Y_{n+1} = \mu + \epsilon_{n+1} + \theta\epsilon_n. \quad (12.44)$$

In the right-hand side of (12.44) 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

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

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.

### 12.12.1 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  $z_{\alpha/2}$  assumes that  $\epsilon_1, \epsilon_2, \dots$  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 Sect. 12.12.2.

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 first consider the errors made when forecasting a stationary AR(1) process. The error in the first prediction is

$$\begin{aligned} Y_{n+1} - \hat{Y}_{n+1} &= \{\mu + \phi(Y_n - \mu) + \epsilon_{n+1}\} - \{\hat{\mu} + \hat{\phi}(Y_n - \hat{\mu})\} \\ &= (\mu - \hat{\mu}) + (\phi - \hat{\phi})Y_n - (\phi\mu - \hat{\phi}\hat{\mu}) + \epsilon_{n+1} \end{aligned} \quad (12.45)$$

$$\approx \epsilon_{n+1}. \quad (12.46)$$

Here (12.45) is the exact error and (12.46) is a “large-sample” approximation. The basis for (12.46) is that as the sample size increases  $\hat{\mu} \rightarrow \mu$  and  $\hat{\phi} \rightarrow \phi$ , so the first three terms in (12.45) 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

$$\begin{aligned} Y_{n+2} - \hat{Y}_{n+2} &= \{\mu + \phi(Y_{n+1} - \mu) + \epsilon_{n+2}\} - \{\mu + \phi(\hat{Y}_{n+1} - \mu)\} \\ &= \phi(Y_{n+1} - \hat{Y}_{n+1}) + \epsilon_{n+2} \\ &= \phi\epsilon_{n+1} + \epsilon_{n+2}. \end{aligned} \quad (12.47)$$

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

$$\begin{aligned} Y_{n+k} - \hat{Y}_{n+k} &\approx \{\mu + \phi(Y_{n+k-1} - \mu) + \epsilon_{n+k}\} - \{\mu + \phi(\hat{Y}_{n+k-1} - \mu)\} \\ &= \phi^{k-1}\epsilon_{n+1} + \phi^{k-2}\epsilon_{n+2} + \cdots + \phi\epsilon_{n+k-1} + \epsilon_{n+k}. \end{aligned} \quad (12.48)$$

By the formula for the sum of a finite geometric series, the variance of the right-hand side of (12.48) is

$$\begin{aligned} \left\{ \phi^{2(k-1)} + \phi^{2(k-2)} + \cdots + \phi^2 + 1 \right\} \sigma_\epsilon^2 &= \left( \frac{1 - \phi^{2k}}{1 - \phi^2} \right) \sigma_\epsilon^2 \\ &\rightarrow \frac{\sigma_\epsilon^2}{1 - \phi^2} \text{ as } k \rightarrow \infty. \end{aligned} \quad (12.49)$$

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 (12.7). 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 (12.49) 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 12.15. Forecasting the one-month inflation rate*

We saw in Example 12.8 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.

Figures 12.18 and 12.19 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.  $\square$

### 12.12.2 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:

$$\begin{aligned} \hat{Y}_{n+t} = & \hat{\mu} + \hat{\phi}_1(\hat{Y}_{n+t-1} - \hat{\mu}) + \cdots + \hat{\phi}_p(\hat{Y}_{n+t-p} - \hat{\mu}) \\ & + \hat{\epsilon}_{n+t} + \hat{\theta}_1\hat{\epsilon}_{n+t-1} + \cdots + \hat{\theta}_q\hat{\epsilon}_{n+t-q}, \quad t = 1, \dots, m, \end{aligned} \quad (12.50)$$



in which

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

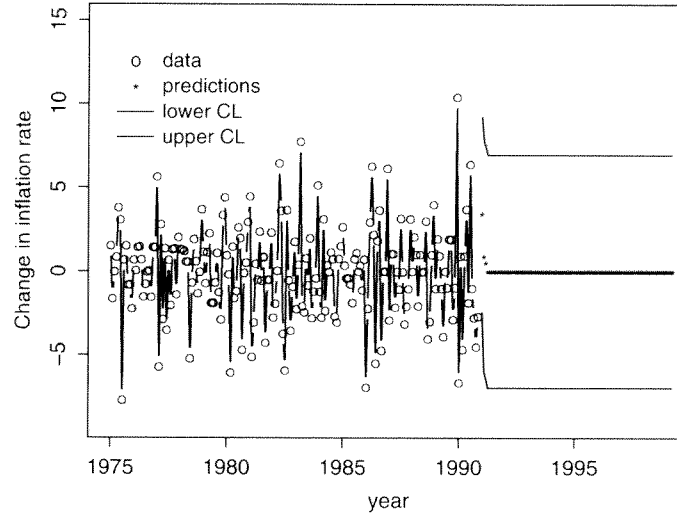


Fig. 12.18. Forecasts of changes in inflation rate.

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 errors generated in step two are mutually independent. For each  $t = 1, \dots, m$ , the  $\alpha/2$ -upper and -lower sample quantiles of the  $B$  random values of  $\hat{Y}_{n+h}$  are the forecast limits for  $Y_{n+h}$ .

To obtain forecasts, rather than forecast limits, one uses  $\hat{\epsilon}_k = 0$ , for  $k = n+1, \dots, n+m$ , in step two. 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,

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

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

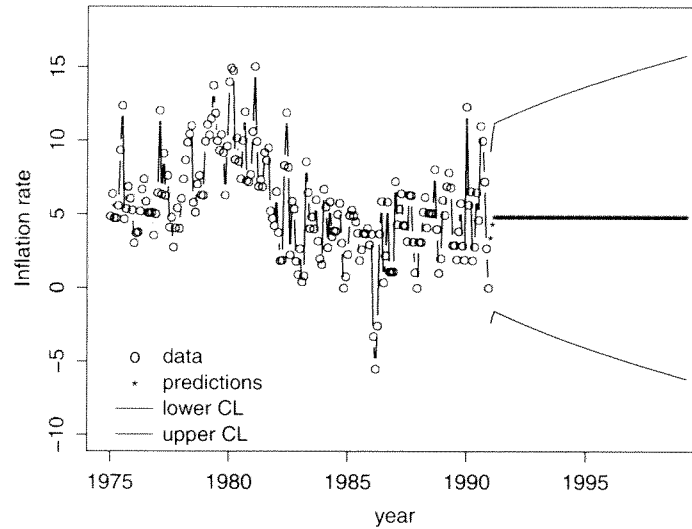
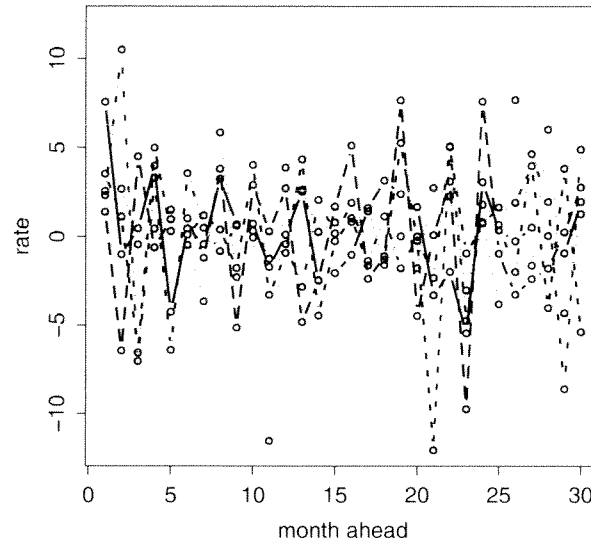


Fig. 12.19. Forecasts of inflation rate.

*Example 12.16. 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 Fig. 12.20. 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 12.21 shows the forecast limits, which are the 2.5% upper and lower sample quantiles. For comparison, the forecast limits generated by R's function `arima()` are also shown. The two sets of forecast limits are very similar even though the `arima()` 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 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 Fig. 12.22. As expected for a nonstationary process, the forecast limits diverge.  $\square$



**Fig. 12.20.** Five random sets of forecasts of changes in the inflation rate computed by simulation.

There are two important advantages to using simulation for forecasting:

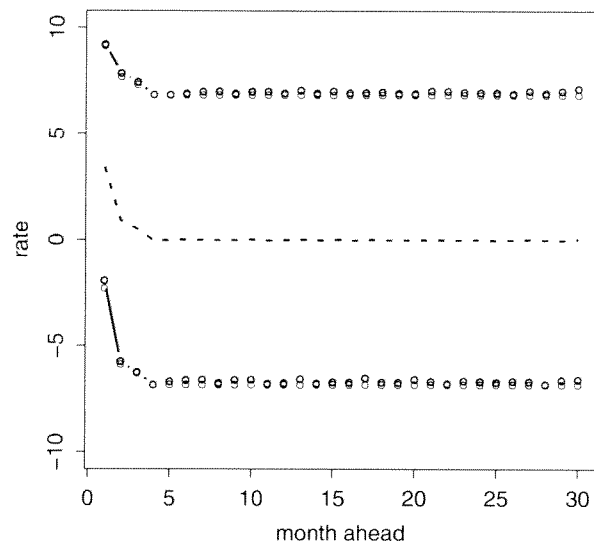
1. simulation can be used in situations where standard software does not compute forecast limits, and
2. simulation does not require that the noise series be Gaussian.

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 than Gaussian tails.

### 12.13 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 on  $Y_{t+1}, \dots, Y_{t+k-1}$ . 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

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



**Fig. 12.21.** 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 visually.

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 12.17. PACF for BMW log returns*

Figure 12.23 is the sample PACF for the BMW log returns computed using the R function `pacf()`. 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 12.4. Note that  $\hat{\phi}_{k,k}$  is outside the test bounds for some values of  $k > 9$ , particularly for  $k = 19$ . This is likely due to random variation.  $\square$

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.

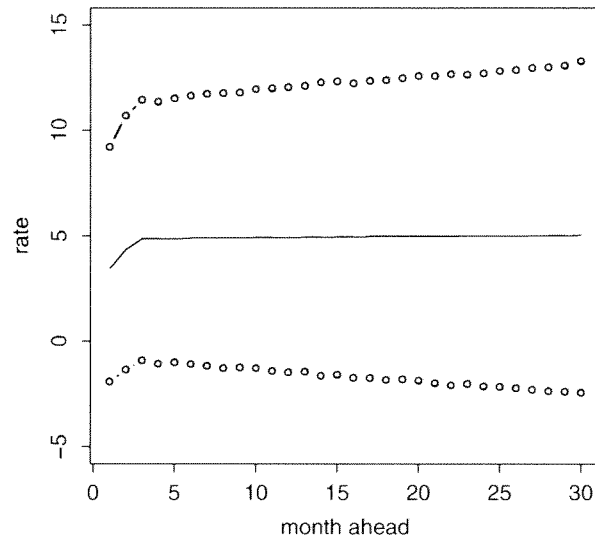


Fig. 12.22. Forecast limits for the inflation rate computed by simulation.

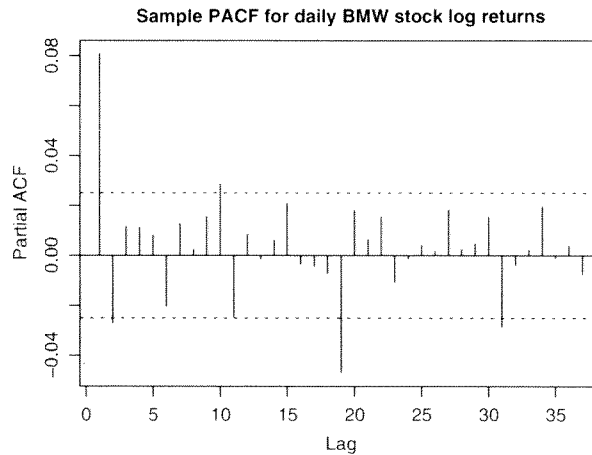


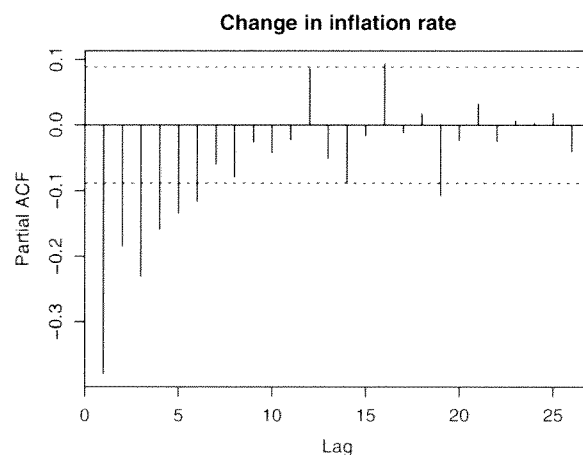
Fig. 12.23. Sample PACF for the daily BMW stock log returns.

*Example 12.18. PACF for changes in the inflation rate*

Figure 12.24 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.  $\square$

## 12.14 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



**Fig. 12.24.** Sample PACF for changes in the inflation rate.

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.

## 12.15 R Lab

### 12.15.1 T-bill Rates

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

```

1 TbGdpPi = read.csv("TbGdpPi.csv", header=TRUE)
2 # r = the 91-day treasury bill rate
3 # y = the log of real GDP
4 # pi = the inflation rate
5 TbGdpPi = ts(TbGdpPi, start = 1955, freq = 4)
6 library(tseries)
7 plot(TbGdpPi)
8 acf(TbGdpPi)
9 adf.test(TbGdpPi[,1])
10 adf.test(TbGdpPi[,2])
11 adf.test(TbGdpPi[,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.

```

12 diff_rate = diff(TbGdpPi)
13 adf.test(diff_rate[,1])
14 adf.test(diff_rate[,2])
15 adf.test(diff_rate[,3])
16 pairs(diff_rate)          # scatterplot matrix
17 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 Chap. 13 and can be ignored for now.

```

18 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 whether 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.

```
19 par(mfrow=c(1,1))
20 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 Sect. 13.1. 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.

```
21 library(forecast)
22 auto.arima(TbGdpPi[,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?*
5. *Carefully express the fitted model chosen by the BIC criterion in mathematical notation.*

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.

```
23 fit1 = arima(TbGdpPi[,1], order=c(?,?,?))
24 acf(residuals(fit1))
25 Box.test(residuals(fit1), lag = 12, type="Ljung", fitdf=?)
```



**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.

```
26 resid2 = (residuals(fit1) - mean(residuals(fit1)))^2
27 acf(resid2)
28 Box.test(resid2, lag = 12, type="Ljung")
```

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

### 12.15.2 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.

```
1 TbGdpPi = read.csv("TbGdpPi.csv", header=TRUE)
2 attach(TbGdpPi)
3 # r = the 91-day treasury bill rate
4 # y = the log of real GDP
5 # pi = the inflation rate
6 # fit the non-seasonal ARIMA model found by auto.arima()
7 # quarterly observations from 1955-1 to 2013-4
8 year = seq(1955,2013.75, by=0.25)
9 library(forecast)
10 auto.arima(pi, max.P=0, max.Q=0, ic="bic")
11 fit = arima(pi, order=c(?,?,?))
12 forecasts = predict(fit, 36)
13 plot(year,pi,xlim=c(1980,2023), ylim=c(-7,12), type="b")
14 lines(seq(from=2014, by=.25, length=36), forecasts$pred, col="red")
15 lines(seq(from=2014, by=.25, length=36),
16       forecasts$pred + 1.96*forecasts$se, col="blue")
17 lines(seq(from=2014, by=.25, length=36),
18       forecasts$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) *Why are the predictions (red) constant throughout?*

## 12.16 Exercises

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

```

1 library(Ecdat)
2 data(CRSPday)
3 crsp=CRSPday[,7]
4 acf(crsp)
5 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(2) models to the CRSP returns:

```

6 arima(crsp,order=c(1,0,0))
7 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 (12.11) to verify Eq. (12.12).  
 9. Show that if  $w_t$  is defined by (12.34) then (12.35) is true.  
 10. For a univariate, discrete time process, what is the difference between a strictly stationary process and a weakly stationary process?  
 11. The time series in the middle and bottom panels of Fig. 12.14 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?  
 12. 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  $Y_{n+2}$ .

13. 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  $Y_{n+2}$ .

14. 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:

```

8 par(mfrow=c(3,2))
9 plot(y,type="l")
10 acf(y)
11 plot(diff(y),type="l")
12 acf(diff(y))
13 plot(diff(y,d=2),type="l")
14 acf(diff(y,d=2))

```

The output was the following figure: