are conducted on individuals, households, business organizations, or other economic entities. The early part of this chapter described regression modeling of cross-section data. **Panel data** typically contain both cross-section and time-series data. These data sets consist of a collection of individuals, households, or corporations that are surveyed repeatedly over time. As an example of a simple econometric model involving time series data, suppose that we wish to develop a model for forecasting monthly consumer spending. A plausible model might be

$$y_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 y_{t-1} + \varepsilon_t,$$

where y_t is consumer spending in month t, x_{t-1} is income in month $t-1$, y_{t-1} is consumer spending in month $t-1$, and ε_t is the random error term. This is a lagged-variable regression model of the type discussed earlier in this chapter.

The consumer spending example above is an example of a very simple single-equation econometric model. Many econometric models involve several equations and the predictor variables in these models can be involved in complex interrelationships with each other and with the dependent variables. For example, consider the following econometric model:

$$\text{Sales} = f_1(\text{GNP, price, number of competitors, advertising expenditures})$$

However, price is likely to be a function of other variables, say

$$\begin{aligned}\text{Price} = f_2(&\text{production costs, distribution costs, overhead costs,}\\ &\text{material cost, packaging costs})\end{aligned}$$

and

$$\begin{aligned}\text{Production costs} &= f_3(\text{production volume, labor costs,}\\ &\qquad\text{material costs, inventory costs})\\ \text{Advertising expenditures} &= f_4(\text{sales, number of competitors})\end{aligned}$$

Notice the interrelationships between these variables. Advertising expenditures certainly influence sales, but the level of sales and the number of competitors will influence the money spent on advertising. Furthermore, different levels of sales will have an impact on production costs.

Constructing and maintaining these models is a complex task. One could (theoretically at least) write a very large number of interrelated equations, but data availability and model estimation issues are practical restrictions. The SAS© software package has good capability for this type of simultaneous equation modeling and is widely used in econometrics. However, forecast accuracy does not necessarily increase with the complexity of the models. Often, a relatively simple time series model will outperform a complex econometric model from a pure forecast accuracy point of view. Econometric models are most useful for providing understanding about the way an economic system works, and for evaluating in a broad sense how different economic policies will perform, and the effect that will have on the economy. This is why their use is largely confined to government entities and some large corporations. There are commercial services that offer econometric models that could be useful to smaller organizations, and free alternatives are available from central banks and other government organizations.

3.10 R COMMANDS FOR CHAPTER 3

Example 3.1 The patient satisfaction data are in the sixth column of the array called patsat.data in which the second and third columns are the age and the severity. Note that we can use the "lm" function to fit the linear model. But as in the example, we will show how to obtain the regression coefficients using the matrix notation.

```
nrow<-dim(patsat.data)[1]
X<-cbind(matrix(1,nrow,1),patsat.data[,2:3])
y<-patsat.data[,6]
beta<-solve(t(X)%*%X)%*%t(X)%*%y
beta
                  [,1]
          143.4720118
 Age       -1.0310534
 Severity  -0.5560378
```

Example 3.3 For this example we will use the "lm" function.

```
satisfaction2.fit<-lm(Satisfaction~Age+Severity+Age:Severity+I(Age^2)+
I(Severity^2), data=patsat)
summary(satisfaction2.fit)
 Call:
 lm(formula = Satisfaction ~ Age+Severity+Age:Severity+I(Age^2)+
     I(Severity^2), data = patsat)

 Residuals:
     Min       1Q  Median       3Q      Max
 -16.915   -3.642   2.015    4.000    9.677
```

```
Coefficients:
                 Estimate Std. Error t value Pr(> |t|)
(Intercept)    127.527542  27.912923   4.569  0.00021 ***
Age             -0.995233   0.702072  -1.418  0.17251
Severity         0.144126   0.922666   0.156  0.87752
I(Age^2)        -0.002830   0.008588  -0.330  0.74534
I(Severity^2)   -0.011368   0.013533  -0.840  0.41134
Age:Severity     0.006457   0.016546   0.390  0.70071
—
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.503 on 19 degrees of freedom
Multiple R-squared:  0.9008,    Adjusted R-squared:  0.8747
F-statistic:  34.5 on 5 and 19 DF,  p-value: 6.76e-09
```

```
anova(satisfaction2.fit)
```

```
Analysis of Variance Table

Response: Satisfaction
              Df Sum Sq Mean Sq  F value     Pr(> F)
Age            1 8756.7  8756.7 155.5644 1.346e-10 ***
Severity       1  907.0   907.0  16.1138 0.0007417 ***
I(Age^2)       1    1.4     1.4   0.0252 0.8756052
I(Severity^2)  1   35.1    35.1   0.6228 0.4397609
Age:Severity   1    8.6     8.6   0.1523 0.7007070
Residuals     19 1069.5    56.3
—
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 3.4 We use the "lm" function again and obtain the linear model. Then we use "confint" function to obtain the confidence intervals of the model parameters. Note that the default confidence level is 95%.

```
satisfaction1.fit<-lm(Satisfaction Age+Severity, data=patsat)
confint(satisfaction1.fit,level=.95)
                 2.5 %      97.5 %
(Intercept) 131.122434 155.8215898
Age          -1.270816  -0.7912905
Severity     -0.828566  -0.2835096
```

Example 3.5 This example refers to the linear model (Satisfaction1.fit). We obtain the confidence and prediction intervals for a new data point for which age = 60 and severity = 60.

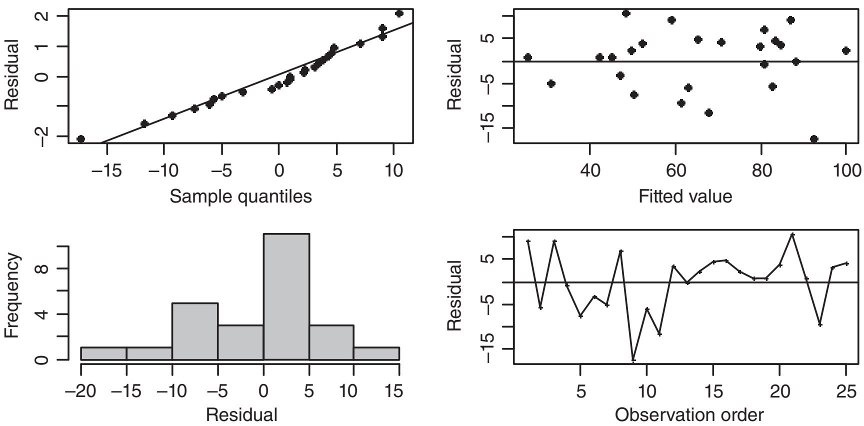
```
new <- data.frame(Age = 60, Severity=60)
pred.sat1.clim<-predict(Satisfaction1.fit, newdata=new, se.fit =
TRUE, interval = "confidence")
pred.sat1.plim<-predict(Satisfaction1.fit, newdata=new, se.fit =
TRUE, interval = "prediction")
pred.sat1.clim$fit
        fit      lwr      upr
 1 48.24654 43.84806 52.64501
pred.sat1.plim$fit
        fit      lwr      upr
 1 48.24654 32.84401 63.64906
```

Example 3.6 We simply repeat Example 3.5 for age = 75 and severity = 60.

```
new <- data.frame(Age = 75, Severity=60)
pred.sat1.clim<-predict(Satisfaction1.fit, newdata=new, se.fit =
TRUE, interval = "confidence")
pred.sat1.plim<-predict(Satisfaction1.fit, newdata=new, se.fit =
TRUE, interval = "prediction")
pred.sat1.clim$fit
        fit      lwr      upr
 1 32.78074 26.99482 38.56666
pred.sat1.plim$fit
        fit      lwr      upr
 1 32.78074 16.92615 48.63533
```

Example 3.7 The residual plots for Satisfaction1.fit can be obtained using the following commands:

```
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(Satisfaction1.fit$res,datax=TRUE,pch=16,xlab=
'Residual',main='')
qqline(Satisfaction1.fit$res,datax=TRUE)
plot(Satisfaction1.fit$fit, Satisfaction1.fit$res,pch=16,
 xlab='Fitted Value',ylab='Residual')
abline(h=0)
hist(Satisfaction1.fit$res,col="gray",xlab='Residual',main='')
plot(Satisfaction1.fit$res,type="l",xlab='Observation
Order',ylab='Residual')
points(Satisfaction1.fit$res,pch=16,cex=.5)
abline(h=0)
```



Example 3.8 In R, one can do stepwise regression using step function which allows for stepwise selection of variables in forward, backward, or both directions. Note that step needs to be applied to a model with 4 input variables as indicated in the example. Therefore we first fit that model and apply the step function. Also note that the variable selection is done based on AIC and therefore we get in the forward selection slightly different results than the one provided in the textbook.

```
satisfaction3.fit<-lm(Satisfaction~Age+Severity+Surg.Med+Anxiety,
data=patsat)

step.for<-step(satisfaction3.fit,direction='forward')

 Start:  AIC=103.18
 Satisfaction~Age + Severity + Surg.Med + Anxiety

step.back<-step(satisfaction3.fit,direction='backward')

 Start:  AIC=103.18
 Satisfaction~Age + Severity + Surg.Med + Anxiety

             Df Sum of Sq    RSS    AIC
 - Surg.Med  1       1.0 1039.9 101.20
 - Anxiety   1      75.4 1114.4 102.93
 <none>                  1038.9 103.18
 - Severity  1     971.5 2010.4 117.68
 - Age       1    3387.7 4426.6 137.41

 Step:  AIC=101.2
 Satisfaction~Age + Severity + Anxiety
```

```
            Df Sum of Sq    RSS    AIC
- Anxiety    1      74.6 1114.5 100.93
<none>                   1039.9 101.20
- Severity   1     971.8 2011.8 115.70
- Age        1    3492.7 4532.6 136.00

Step:  AIC=100.93
Satisfaction ~ Age + Severity

            Df Sum of Sq    RSS    AIC
<none>                   1114.5 100.93
- Severity   1     907.0 2021.6 113.82
- Age        1    4029.4 5143.9 137.17
```

```
step.both<-step(satisfaction3.fit,direction='both')
```

```
Start:  AIC=103.18
Satisfaction ~ Age + Severity + Surg.Med + Anxiety

            Df Sum of Sq    RSS    AIC
- Surg.Med   1       1.0 1039.9 101.20
- Anxiety    1      75.4 1114.4 102.93
<none>                   1038.9 103.18
- Severity   1     971.5 2010.4 117.68
- Age        1    3387.7 4426.6 137.41

Step:  AIC=101.2
Satisfaction ~ Age + Severity + Anxiety

            Df Sum of Sq    RSS    AIC
- Anxiety    1      74.6 1114.5 100.93
<none>                   1039.9 101.20
+ Surg.Med   1       1.0 1038.9 103.18
- Severity   1     971.8 2011.8 115.70
- Age        1    3492.7 4532.6 136.00

Step:  AIC=100.93
Satisfaction ~ Age + Severity

            Df Sum of Sq    RSS    AIC
<none>                   1114.5 100.93
+ Anxiety    1      74.6 1039.9 101.20
+ Surg.Med   1       0.2 1114.4 102.93
- Severity   1     907.0 2021.6 113.82
- Age        1    4029.4 5143.9 137.17
```

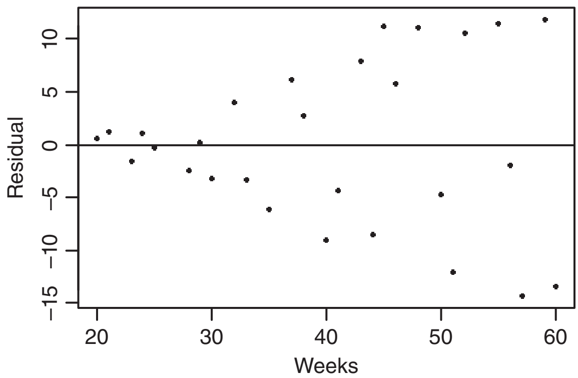
In R, one can do best subset regression using leaps function from leaps package. We first upload the leaps package.

```
library(leaps)
```

```
step.best<-regsubsets(Satisfaction~Age+Severity+Surg.Med+Anxiety,
data=patsat)
summary(step.best)
 Subset selection object
 Call: regsubsets.formula(Satisfaction~Age+Severity+Surg.Med+
     Anxiety, data = patsat)
 4 Variables  (and intercept)
          Forced in Forced out
 Age          FALSE      FALSE
 Severity     FALSE      FALSE
 Surg.Med     FALSE      FALSE
 Anxiety      FALSE      FALSE
 1 subsets of each size up to 4
 Selection Algorithm: exhaustive
          Age Severity Surg.Med Anxiety
 1  ( 1 ) "*" " "      " "      " "
 2  ( 1 ) "*" "*"      " "      " "
 3  ( 1 ) "*" "*"      " "      "*"
 4  ( 1 ) "*" "*"      "*"      "*"
```

Example 3.9 The connector strength data are in the second column of the array called strength.data in which the first column is the Weeks. We start with fitting the linear model and plot the residuals vs. Weeks.

```
strength1.fit<-lm(Strength~Weeks, data=strength.data)
summary(strength1.fit)
 Call:
 lm(formula = Strength~Weeks, data = strength.data)
 Residuals:
      Min       1Q   Median       3Q      Max
 -14.3639  -4.4449  -0.0861   5.8671  11.8842

 Coefficients:
             Estimate Std. Error t value Pr(> |t|)
 (Intercept)  25.9360     5.1116   5.074 2.76e-05 ***
 Weeks         0.3759     0.1221   3.078  0.00486 **
 —
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 7.814 on 26 degrees of freedom
 Multiple R-squared:  0.2671,    Adjusted R-squared:  0.2389
 F-statistic: 9.476 on 1 and 26 DF,  p-value: 0.004863

plot(strength.data[,1],strength1.fit$res, pch=16,cex=.5,
xlab='Weeks',ylab='Residual')
abline(h=0)
```



We then fit a linear model to the absolute value of the residuals and obtain the weights as the inverse of the square of the fitted values.

```
res.fit<-lm(abs(strength1.fit$res) ~ Weeks, data=strength.data)
weights.strength<-1/(res.fit$fitted^2)
```

We then fit a linear model to the absolute value of the residuals and obtain the weights as the inverse of the square of the fitted values.

```
strength2.fit<-lm(Strength ~ Weeks, data=strength.data,
weights=weights.strength)
summary(strength2.fit)

 Call:
 lm(formula = Strength ~ Weeks, data = strength.data, weights =
weights.strength)

 Weighted Residuals:
     Min      1Q  Median      3Q     Max
 -1.9695 -1.0450 -0.1231  1.1507  1.5785

 Coefficients:
             Estimate Std. Error t value Pr(> |t|)
 (Intercept) 27.54467    1.38051  19.953 <  2e-16 ***
 Weeks        0.32383    0.06769   4.784 5.94e-05 ***
 —
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 1.119 on 26 degrees of freedom
 Multiple R-squared:  0.4682,    Adjusted R-squared:  0.4477
 F-statistic: 22.89 on 1 and 26 DF,  p-value: 5.942e-05
```

Example 3.12 Different packages such as car, lmtest, and bstats offer functions for Durbin–Watson test. We will use function "dwt" in package car. Note that dwt function allows for two-sided or one-sided tests. As in the example we will test for positive autocorrelation. The data are given in softsales.data where the columns are Year, Sales, Expenditures and Population (to be used in the next example).

```
library(car)
soft1.fit<-lm(Sales~Expenditures, data=softsales.data)
dwt(soft1.fit, alternative="positive")
     lag Autocorrelation D-W Statistic p-value
        1       0.3354445       1.08005   0.007
      Alternative hypothesis: rho >  0
```

Since the *p*-value is too small the null hypothesis is rejected concluding that the errors are positively correlated.

Example 3.13 We repeat Example 3.12 with the model expanded to include Population as well.

```
soft2.fit<-lm(Sales~Expenditures+Population, data=softsales.data)
dwt(soft2.fit, alternative="positive")
     lag Autocorrelation D-W Statistic p-value
        1       -0.534382      3.059322   0.974
      Alternative hypothesis: rho >  0
```

As concluded in the example, adding the input variable Population seems to resolve the autocorrelation issue resulting in large *p*-value for the test for autocorrelation.

Example 3.14 The Cochrane–Orcutt method can be found in package orcutt. The function to be used is "cochran.orcutt". The data are given in toothsales.data where the columns are Share and Price.

```
library(orcutt)
tooth1.fit<-lm(Share~Price, data=toothsales.data)
dwt(tooth1.fit, alternative="positive")
          lag Autocorrelation D-W Statistic p-value
        1       0.4094368      1.135816   0.005
      Alternative hypothesis: rho >  0
tooth2.fit<-cochrane.orcutt(tooth1.fit)
```

```
tooth2.fit
 $Cochrane.Orcutt

 Call:
 lm(formula = YB ~ XB - 1)

 Residuals:
      Min       1Q   Median       3Q      Max
 -0.55508 -0.25069 -0.05506  0.25007  0.83017

 Coefficients:
               Estimate Std. Error t value Pr(> |t|)
 XB(Intercept)   26.722      1.633   16.36 7.71e-12 ***
 XBPrice        -24.084      1.938  -12.42 5.90e-10 ***
 —
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 0.3955 on 17 degrees of freedom
 Multiple R-squared:  0.991,     Adjusted R-squared:  0.9899
 F-statistic: 932.7 on 2 and 17 DF,  p-value: <  2.2e-16

 $rho
 [1] 0.4252321
 $number.interaction
 [1] 8
```

The results are not exactly the same as the ones given in the example. This should be due to the difference in the approaches. Where the book uses a two-step procedure, the function Cochrane.Orcutt "estimates both autocorrelation and beta coefficients recursively until we reach the convergence (8th decimal)".

Example 3.15 For this example we will use "gls" function in package nlme. From Example 3.14 we know that there is autocorrelation in the residuals of the linear model. We first assume that the first-order model will be sufficient to model the autocorrelation as shown in the book.

```
tooth3.fit <- gls(Share ~ Price, data = toothsales.data,
correlation=corARMA(p=1), method="ML")
summary(tooth3.fit)
 Generalized least squares fit by maximum likelihood
   Model: Share ~ Price
   Data: toothsales.data
        AIC      BIC    logLik
   24.90475 28.88767 -8.452373

 Correlation Structure: AR(1)
  Formula: ~ 1
```

```
  Parameter estimate(s):
       Phi
 0.4325871

 Coefficients:
                 Value Std.Error   t-value p-value
 (Intercept)  26.33218  1.436305  18.33328       0
 Price       -23.59030  1.673740 -14.09436       0
  Correlation:
       (Intr)
 Price -0.995

 Standardized residuals:
        Min          Q1         Med          Q3         Max
 -1.85194806 -0.85848738  0.08945623  0.69587678  2.03734437

 Residual standard error: 0.4074217
 Degrees of freedom: 20 total; 18 residual

intervals(tooth3.fit)

 Approximate 95% confidence intervals

  Coefficients:
                 lower      est.     upper
 (Intercept)  23.31462  26.33218  29.34974
 Price       -27.10670 -23.59030 -20.07390
 attr(,"label")

 [1] "Coefficients:"

  Correlation structure:
           lower      est.     upper
 Phi -0.04226294 0.4325871 0.7480172
 attr(,"label")

 [1] "Correlation structure:"

  Residual standard error:
     lower      est.     upper
 0.2805616 0.4074217 0.5916436

predict(tooth3.fit)
```

The second-order autoregressive model for the errors can be fitted using,

```
tooth4.fit <- gls(Share ~ Price, data = toothsales.data,
correlation=corARMA(p=2), method="ML")
```

Example 3.16 To create the lagged version of the variables and also adjust for the number of observations, we use the following commands:

```
T<-length(toothsales.data$Share)
yt<-toothsales.data$Share[2:T]
yt.lag1<- toothsales.data$Share[1:(T-1)]
xt<-toothsales.data$Price[2:T]
xt.lag1<- toothsales.data$Price[1:(T-1)]
tooth5.fit<-lm(yt ~ yt.lag1+xt+xt.lag1)
summary(tooth5.fit)
 Call:
 lm(formula = yt ~ yt.lag1 + xt + xt.lag1)
 Residuals:
      Min        1Q    Median        3Q       Max
 -0.59958 -0.23973 -0.02918  0.26351  0.66532

 Coefficients:
             Estimate Std. Error t value Pr(> |t|)
 (Intercept)  16.0675     6.0904   2.638   0.0186 *
 yt.lag1       0.4266     0.2237   1.907   0.0759 .
 xt          -22.2532     2.4870  -8.948 2.11e-07 ***
 xt.lag1       7.5941     5.8697   1.294   0.2153
 —
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Residual standard error: 0.402 on 15 degrees of freedom
 Multiple R-squared:   0.96,     Adjusted R-squared:  0.952
 F-statistic: 120.1 on 3 and 15 DF,  p-value: 1.037e-10
```

EXERCISES

3.1 An article in the journal *Air and Waste* (Update on Ozone Trends in California's South Coast Air Basin, Vol. 43, 1993) investigated the ozone levels in the South Coast Air Basin of California for the years 1976–1991. The author believes that the number of days the ozone levels exceeded 0.20 ppm (the response) depends on the seasonal meteorological index, which is the seasonal average 850-millibar Temperature (the predictor). Table E3.1 gives the data.

- **a.** Construct a scatter diagram of the data.
- **b.** Estimate the prediction equation.
- **c.** Test for significance of regression.
- **d.** Calculate the 95% CI and PI on for a seasonal meteorological index value of 17. Interpret these quantities.
- **e.** Analyze the residuals. Is there evidence of model inadequacy?
- **f.** Is there any evidence of autocorrelation in the residuals?

TABLE E3.1 Days that Ozone Levels Exceed 20 ppm and Seasonal Meteorological Index

Year	Days	Index
1976	91	16.7
1977	105	17.1
1978	106	18.2
1979	108	18.1
1980	88	17.2
1981	91	18.2
1982	58	16.0
1983	82	17.2
1984	81	18.0
1985	65	17.2
1986	61	16.9
1987	48	17.1
1988	61	18.2
1989	43	17.3
1990	33	17.5
1991	36	16.6

3.2 Montgomery, Peck, and Vining (2012) present data on the number of pounds of steam used per month at a plant. Steam usage is thought to be related to the average monthly ambient temperature. The past year's usages and temperatures are shown in Table E3.2.

TABLE E3.2 Monthly Steam Usage and Average Ambient Temperature

Month	Temperature (°F)	Usage/1000	Month	Temperature (°F)	Usage/1000
January	21	185.79	July	68	621.55
February	24	214.47	August	74	675.06
March	32	288.03	September	62	562.03
April	47	424.84	October	50	452.93
May	50	454.68	November	41	369.95
June	59	539.03	December	30	273.98

a. Fit a simple linear regression model to the data.

b. Test for significance of regression.

c. Analyze the residuals from this model.

d. Plant management believes that an increase in average ambient temperature of one degree will increase average monthly steam consumption by 10,000 lb. Do the data support this statement?

e. Construct a 99% prediction interval on steam usage in a month with average ambient temperature of 58°F.

3.3 On March 1, 1984, the *Wall Street Journal* published a survey of television advertisements conducted by Video Board Tests, Inc., a New York ad-testing company that interviewed 4000 adults. These people were regular product users who were asked to cite a commercial they had seen for that product category in the past week. In this case, the response is the number of millions of retained impressions per week. The predictor variable is the amount of money spent by the firm on advertising. The data are in Table E3.3.

TABLE E3.3 Number of Retained Impressions and Advertising Expenditures

Firm	Amount Spent (Millions)	Retained Impressions per Week (Millions)
Miller Lite	50.1	32.1
Pepsi	74.1	99.6
Stroh's	19.3	11.7
Federal Express	22.9	21.9
Burger King	82.4	60.8
Coca-Cola	40.1	78.6
McDonald's	185.9	92.4
MCI	26.9	50.7
Diet Cola	20.4	21.4
Ford	166.2	40.1
Levi's	27	40.8
Bud Lite	45.6	10.4
ATT Bell	154.9	88.9
Calvin Klein	5	12
Wendy's	49.7	29.2
Polaroid	26.9	38
Shasta	5.7	10
Meow Mix	7.6	12.3
Oscar Meyer	9.2	23.4
Crest	32.4	71.1
Kibbles N Bits	6.1	4.4

a. Fit the simple linear regression model to these data.

b. Is there a significant relationship between the amount that a company spends on advertising and retained impressions? Justify your answer statistically.

c. Analyze the residuals from this model.

d. Construct the 95% confidence intervals on the regression coefficients.

e. Give the 95% confidence and prediction intervals for the number of retained impressions for MCI.

3.4 Suppose that we have fit the straight-line regression model $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$, but the response is affected by a second variable x_2 such that the true regression function is

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

a. Is the least squares estimator of the slope in the original simple linear regression model unbiased?

b. Show the bias in $\hat{\beta}_1$.

3.5 Suppose that we are fitting a straight line and wish to make the standard error of the slope as small as possible. Suppose that the "region of interest" for x is $-1 \leq x \leq 1$. Where should the observations $x_1, x_2, \ldots, x_n$ be taken? Discuss the practical aspects of this data collection plan.

3.6 Consider the simple linear regression model

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

where the intercept β_0 is known.

a. Find the least squares estimator of β_1 for this model. Does this answer seem reasonable?

b. What is the variance of the slope $(\hat{\beta}_1)$ for the least squares estimator found in part a?

c. Find a $100(1 - \alpha)$ percent CI for β_1. Is this interval narrower than the estimator for the case where both slope and intercept are unknown?

3.7 The quality of Pinot Noir wine is thought to be related to the properties of clarity, aroma, body, flavor, and oakiness. Data for 38 wines are given in Table E3.4.

TABLE E3.4 Wine Quality Data[a] (Found in Minitab)

Clarity, x_1	Aroma, x_2	Body, x_3	Flavor, x_4	Oakiness, x_5	Quality, y	Region
1	3.3	2.8	3.1	4.1	9.8	1
1	4.4	4.9	3.5	3.9	12.6	1
1	3.9	5.3	4.8	4.7	11.9	1
1	3.9	2.6	3.1	3.6	11.1	1
1	5.6	5.1	5.5	5.1	13.3	1
1	4.6	4.7	5	4.1	12.8	1
1	4.8	4.8	4.8	3.3	12.8	1
1	5.3	4.5	4.3	5.2	12	1
1	4.3	4.3	3.9	2.9	13.6	3
1	4.3	3.9	4.7	3.9	13.9	1
1	5.1	4.3	4.5	3.6	14.4	3
0.5	3.3	5.4	4.3	3.6	12.3	2
0.8	5.9	5.7	7	4.1	16.1	3
0.7	7.7	6.6	6.7	3.7	16.1	3
1	7.1	4.4	5.8	4.1	15.5	3
0.9	5.5	5.6	5.6	4.4	15.5	3
1	6.3	5.4	4.8	4.6	13.8	3
1	5	5.5	5.5	4.1	13.8	3
1	4.6	4.1	4.3	3.1	11.3	1
0.9	3.4	5	3.4	3.4	7.9	2
0.9	6.4	5.4	6.6	4.8	15.1	3
1	5.5	5.3	5.3	3.8	13.5	3
0.7	4.7	4.1	5	3.7	10.8	2
0.7	4.1	4	4.1	4	9.5	2
1	6	5.4	5.7	4.7	12.7	3
1	4.3	4.6	4.7	4.9	11.6	2
1	3.9	4	5.1	5.1	11.7	1
1	5.1	4.9	5	5.1	11.9	2
1	3.9	4.4	5	4.4	10.8	2
1	4.5	3.7	2.9	3.9	8.5	2
1	5.2	4.3	5	6	10.7	2
0.8	4.2	3.8	3	4.7	9.1	1
1	3.3	3.5	4.3	4.5	12.1	1
1	6.8	5	6	5.2	14.9	3
0.8	5	5.7	5.5	4.8	13.5	1
0.8	3.5	4.7	4.2	3.3	12.2	1
0.8	4.3	5.5	3.5	5.8	10.3	1
0.8	5.2	4.8	5.7	3.5	13.2	1

[a] The wine here is Pinot Noir. Region refers to distinct geographic regions.

a. Fit a multiple linear regression model relating wine quality to these predictors. Do not include the "Region" variable in the model.

b. Test for significance of regression. What conclusions can you draw?

c. Use t-tests to assess the contribution of each predictor to the model. Discuss your findings.

d. Analyze the residuals from this model. Is the model adequate?

e. Calculate R^2 and the adjusted R^2 for this model. Compare these values to the R^2 and adjusted R^2 for the linear regression model relating wine quality to only the predictors "Aroma" and "Flavor." Discuss your results.

f. Find a 95% CI for the regression coefficient for "Flavor" for both models in part e. Discuss any differences.

3.8 Reconsider the wine quality data in Table E3.4. The "Region" predictor refers to three distinct geographical regions where the wine was produced. Note that this is a categorical variable.

a. Fit the model using the "Region" variable as it is given in Table E3.4. What potential difficulties could be introduced by including this variable in the regression model using the three levels shown in Table E3.4?

b. An alternative way to include the categorical variable "Region" would be to introduce two indicator variables x_1 and x_2 as follows:

Region	x_1	x_2
1	0	0
2	1	0
3	0	1

Why is this approach better than just using the codes 1, 2, and 3?

c. Rework Exercise 3.7 using the indicator variables defined in part b for "Region."

3.9 Table B.6 in Appendix B contains data on the global mean surface air temperature anomaly and the global CO_2 concentration. Fit a regression model to these data, using the global CO_2 concentration as the predictor. Analyze the residuals from this model. Is there evidence of autocorrelation in these data? If so, use one iteration of the Cochrane–Orcutt method to estimate the parameters.

3.10 Table B.13 in Appendix B contains hourly yield measurements from a chemical process and the process operating temperature. Fit a regression model to these data, using the temperature as the predictor. Analyze the residuals from this model. Is there evidence of autocorrelation in these data?

3.11 The data in Table E3.5 give the percentage share of market of a particular brand of canned peaches (y_t) for the past 15 months and the relative selling price (x_t).

TABLE E3.5 Market Share and Price of Canned Peaches

t	x_t	y_t	t	x_t	y_t
1	100	15.93	9	85	16.60
2	98	16.26	10	83	17.16
3	100	15.94	11	81	17.77
4	89	16.81	12	79	18.05
5	95	15.67	13	90	16.78
6	87	16.47	14	77	18.17
7	93	15.66	15	78	17.25
8	82	16.94			

a. Fit a simple linear regression model to these data. Plot the residuals versus time. Is there any indication of autocorrelation?

b. Use the Durbin–Watson test to determine if there is positive autocorrelation in the errors. What are your conclusions?

c. Use one iteration of the Cochrane–Orcutt procedure to estimate the regression coefficients. Find the standard errors of these regression coefficients.

d. Is there positive autocorrelation remaining after the first iteration? Would you conclude that the iterative parameter estimation technique has been successful?

3.12 The data in Table E3.6 give the monthly sales for a cosmetics manufacturer (y_t) and the corresponding monthly sales for the entire industry (x_t). The units of both variables are millions of dollars.

a. Build a simple linear regression model relating company sales to industry sales. Plot the residuals against time. Is there any indication of autocorrelation?

TABLE E3.6 Cosmetic Sales Data for Exercise 3.12

t	x_t	y_t	t	x_t	y_t
1	5.00	0.318	10	6.16	0.650
2	5.06	0.330	11	6.22	0.655
3	5.12	0.356	12	6.31	0.713
4	5.10	0.334	13	6.38	0.724
5	5.35	0.386	14	6.54	0.775
6	5.57	0.455	15	6.68	0.78
7	5.61	0.460	16	6.73	0.796
8	5.80	0.527	17	6.89	0.859
9	6.04	0.598	18	6.97	0.88

b. Use the Durbin–Watson test to determine if there is positive autocorrelation in the errors. What are your conclusions?

c. Use one iteration of the Cochrane–Orcutt procedure to estimate the model parameters. Compare the standard error of these regression coefficients with the standard error of the least squares estimates.

d. Test for positive autocorrelation following the first iteration. Has the procedure been successful?

3.13 Reconsider the data in Exercise 3.12. Define a new set of transformed variables as the first difference of the original variables, $y'_t = y_t - y_{t-1}$ and $x'_t = x_t - x_{t-1}$. Regress y'_t on x'_t through the origin. Compare the estimate of the slope from this first-difference approach with the estimate obtained from the iterative method in Exercise 3.12.

3.14 Show that an equivalent way to perform the test for significance of regression in multiple linear regression is to base the test on R^2 as follows. To test $H_0 : \beta_1 = \beta_2 = \cdots = \beta_k$ versus H_1: at least one $\beta_j \neq 0$, calculate

$$F_0 = \frac{R^2(n-p)}{k(1-R^2)}$$

and reject H_0 if the computed value of F_0 exceeds $F_{a,k,n-p}$, where $p = k + 1$.

3.15 Suppose that a linear regression model with $k = 2$ regressors has been fit to $n = 25$ observations and $R^2 = 0.90$.

a. Test for significance of regression at $\alpha = 0.05$. Use the results of the Exercise 3.14.

b. What is the smallest value of R^2 that would lead to the conclusion of a significant regression if $\alpha = 0.05$? Are you surprised at how small this value of R^2 is?

3.16 Consider the simple linear regression model $y_t = \beta_0 + \beta_1 x + \varepsilon_t$, where the errors are generated by the second-order autoregressive process

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + a_t$$

Discuss how the Cochrane–Orcutt iterative procedure could be used in this situation. What transformations would be used on the variables y_t and x_t? How would you estimate the parameters ρ_1 and ρ_2?

3.17 Show that an alternate computing formula for the regression sum of squares in a linear regression model is

$$SS_{\mathrm{R}} = \sum_{i=1}^{n} \hat{y}_i^2 - n\bar{y}^2$$

3.18 An article in *Quality Engineering* (The Catapult Problem: Enhanced Engineering Modeling Using Experimental Design, Vol. 4, 1992) conducted an experiment with a catapult to determine the effects of hook (x_1), arm length (x_2), start angle (x_3), and stop angle (x_4) on the distance that the catapult throws a ball. They threw the ball three times for each setting of the factors. Table E3.7 summarizes the experimental results.

TABLE E3.7 Catapult Experiment Data for Exercise 3.18

x_1	x_2	x_3	x_4		y	
−1	−1	−1	−1	28.0	27.1	26.2
−1	−1	1	1	46.5	43.5	46.5
−1	1	−1	1	21.9	21.0	20.1
−1	1	1	−1	52.9	53.7	52.0
1	−1	−1	1	75.0	73.1	74.3
1	−1	1	−1	127.7	126.9	128.7
1	1	−1	−1	86.2	86.5	87.0
1	1	1	1	195.0	195.9	195.7

a. Fit a regression model to the data and perform a residual analysis for the model.

b. Use the sample variances as the basis for WLS estimation of the original data (not the sample means).

c. Fit an appropriate model to the sample variances. Use this model to develop the appropriate weights and repeat part b.

3.19 Consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where the variance of ε_i is proportional to x_i^2; that is, $\text{Var}(\varepsilon_i) = \sigma^2 x_i^2$.

a. Suppose that we use the transformations $y' = y/x$ and $x' = 1/x$. Is this a variance-stabilizing transformation?

b. What are the relationships between the parameters in the original and transformed models?

c. Suppose we use the method of WLS with $w_i = 1/x_i^2$. Is this equivalent to the transformation introduced in part a?

3.20 Consider the WLS normal equations for the case of simple linear regression where time is the predictor variable, Eq. (3.62). Suppose that the variances of the errors are proportional to the index of time such that $w_t = 1/t$. Simplify the normal equations for this situation. Solve for the estimates of the model parameters.

3.21 Consider the simple linear regression model where time is the predictor variable. Assume that the errors are uncorrelated and have constant variance σ^2. Show that the variances of the model parameter estimates are

$$V(\hat{\beta}_0) = \sigma^2 \frac{2(2T+1)}{T(T-1)}$$

and

$$V(\hat{\beta}_1) = \sigma^2 \frac{12}{T(T^2-1)}$$

3.22 Analyze the regression model in Exercise 3.1 for leverage and influence. Discuss your results.

3.23 Analyze the regression model in Exercise 3.2 for leverage and influence. Discuss your results.

3.24 Analyze the regression model in Exercise 3.3 for leverage and influence. Discuss your results.

3.25 Analyze the regression model for the wine quality data in Exercise 3.7 for leverage and influence. Discuss your results.

3.26 Consider the wine quality data in Exercise 3.7. Use variable selection techniques to determine an appropriate regression model for these data.

3.27 Consider the catapult data in Exercise 3.18. Use variable selection techniques to determine an appropriate regression model for these data. In determining the candidate variables, consider all of the two-factor cross-products of the original four variables.

3.28 Table B.10 in Appendix B presents monthly data on airline miles flown in the United Kingdom. Fit an appropriate regression model to these data. Analyze the residuals and comment on model adequacy.

3.29 Table B.11 in Appendix B presents data on monthly champagne sales. Fit an appropriate regression model to these data. Analyze the residuals and comment on model adequacy.

3.30 Consider the data in Table E3.5. Fit a time series regression model with autocorrected errors to these data. Compare this model with the results you obtained in Exercise 3.11 using the Cochrane–Orcutt procedure.

3.31 Consider the data in Table E3.5. Fit the lagged variables regression models shown in Eqs. (3.119) and (3.120) to these data. Compare these models with the results you obtained in Exercise 3.11 using the Cochrane–Orcutt procedure, and with the time series regression model from Exercise 3.30.

3.32 Consider the data in Table E3.5. Fit a time series regression model with autocorrected errors to these data. Compare this model with the results you obtained in Exercise 3.13 using the Cochrane–Orcutt procedure.

3.33 Consider the data in Table E3.6. Fit the lagged variables regression models shown in Eqs. (3.119) and (3.120) to these data. Compare these models with the results you obtained in Exercise 3.13 using the Cochrane–Orcutt procedure, and with the time series regression model from Exercise 3.32.

3.34 Consider the global surface air temperature anomaly data and the CO_2 concentration data in Table B.6 in Appendix B. Fit a time series regression model to these data, using global surface air temperature anomaly as the response variable. Is there any indication of autocorrelation in the residuals? What corrective action and modeling strategies would you recommend?

3.35 Table B.20 in Appendix B contains data on tax refund amounts and population. Fit an OLS regression model to these data.

a. Analyze the residuals and comment on model adequacy.

b. Fit the lagged variables regression models shown in Eqs. (3.119) and (3.120) to these data. How do these models compare with the OLS model in part a?

3.36 Table B.25 contains data from the National Highway Traffic Safety Administration on motor vehicle fatalities from 1966 to 2012, along with several other variables. These data are used by a variety of governmental and industry groups, as well as research organizations.

a. Plot the fatalities data. Comment on the graph.

b. Construct a scatter plot of fatalities versus number of licensed drivers. Comment on the apparent relationship between these two factors.

c. Fit a simple linear regression model to the fatalities data, using the number of licensed drivers as the predictor variable. Discuss the summary statistics from this model.

d. Analyze the residuals from the model in part c. Discuss the adequacy of the fitted model.

e. Calculate the Durbin–Watson test statistic for the model in part c. Is there evidence of autocorrelation in the residuals? Is a time series regression model more appropriate than an OLS model for these data?

3.37 Consider the motor vehicle fatalities data in Appendix Table B.25 and the simple linear regression model from Exercise 3.36. There are several candidate predictors that could be added to the model. Add the number of registered motor vehicles to the model that you fit in Exercise 3.36. Has the addition of another predictor improved the model?

3.38 Consider the motor vehicle fatalities data in Appendix Table B.25. There are several candidate predictors that could be added to the model. Use stepwise regression to find an appropriate subset of predictors for the fatalities data. Analyze the residuals from the model, including the Durbin–Watson test, and comment on model adequacy.

3.39 Consider the motor vehicle fatalities data in Appendix Table B.25. There are several candidate predictors that could be added to the model. Use an all-possible-models approach to find an appropriate subset of predictors for the fatalities data. Analyze the residuals from the model, including the Durbin–Watson test, and comment on model adequacy. Compare this model to the one you obtained through stepwise model fitting in Exercise 3.38.

3.40 Appendix Table B.26 contains data on monthly single-family residential new home sales from 1963 through 2014. The number of building permits issued is also given in the table.

a. Plot the home sales data. Comment on the graph.

b. Construct a scatter plot of home sales versus number of building permits. Comment on the apparent relationship between these two factors.

c. Fit a simple linear regression model to the home sales data, using the number of building permits as the predictor variable. Discuss the summary statistics from this model.

d. Analyze the residuals from the model in part c. Discuss the adequacy of the fitted model.

e. Calculate the Durbin–Watson test statistic for the model in part c. Is there evidence of autocorrelation in the residuals? Is a time series regression model more appropriate than an OLS model for these data?

CHAPTER 4

EXPONENTIAL SMOOTHING METHODS

If you have to forecast, forecast often.

EDGAR R. FIEDLER, *American economist*

4.1 INTRODUCTION

We can often think of a data set as consisting of two distinct components: **signal** and **noise**. Signal represents any pattern caused by the intrinsic dynamics of the process from which the data are collected. These patterns can take various forms from a simple constant process to a more complicated structure that cannot be extracted visually or with any basic statistical tools. The constant process, for example, is represented as

$$y_t = \mu + \varepsilon_t, \tag{4.1}$$

where μ represents the underlying constant level of system response and ε_t is the noise at time t. The ε_t is often assumed to be uncorrelated with mean 0 and constant variance σ_ϵ^2.

We have already discussed some basic data smoothers in Section 2.2.2. **Smoothing** can be seen as a technique to separate the signal and the noise

Introduction to Time Series Analysis and Forecasting, Second Edition.
Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci.

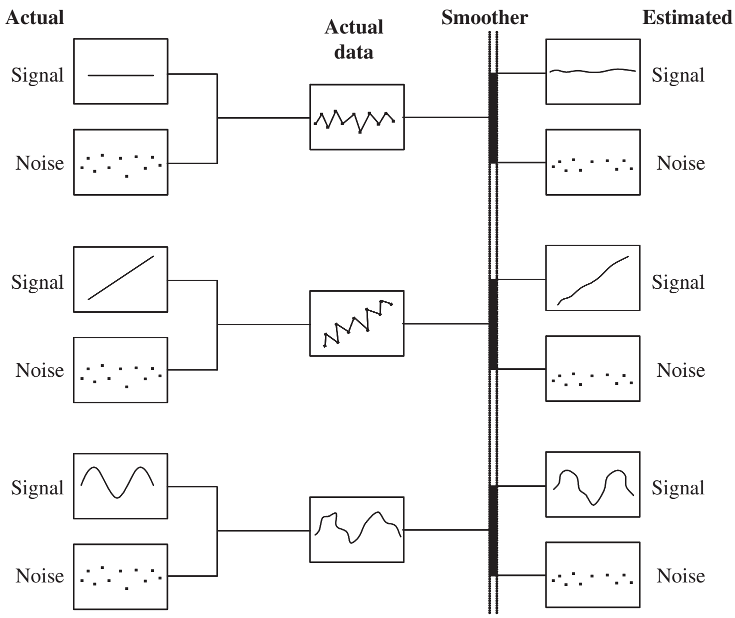


FIGURE 4.1 The process of smoothing a data set.

as much as possible and in that a smoother acts as a filter to obtain an "estimate" for the signal. In Figure 4.1, we give various types of signals that with the help of a smoother can be "reconstructed" and the underlying pattern of the signal is to some extent recovered. The smoothers that we will discuss in this chapter achieve this by simply relating the current observation to the previous ones. For a given data set, one can devise forward and/or backward looking smoothers but in this chapter we will only consider backward looking smoothers. That is, at any given T, the observation y_T will be replaced by a combination of observations at and before T. It does then intuitively make sense to use some sort of an "average" of the current and the previous observations to smooth the data. An obvious choice is to replace the current observation with the average of the observations at $T, T-1, \ldots, 1$. In fact this is the "best" choice in the least squares sense for a constant process given in Eq. (4.1).

A constant process can be smoothed by replacing the current observation with the best estimate for μ. Using the least squares criterion, we define the error sum of squares, $\mathrm{SS_E}$, for the constant process as

$$SS_E = \sum_{t=1}^{T} (y_t - \mu)^2.$$

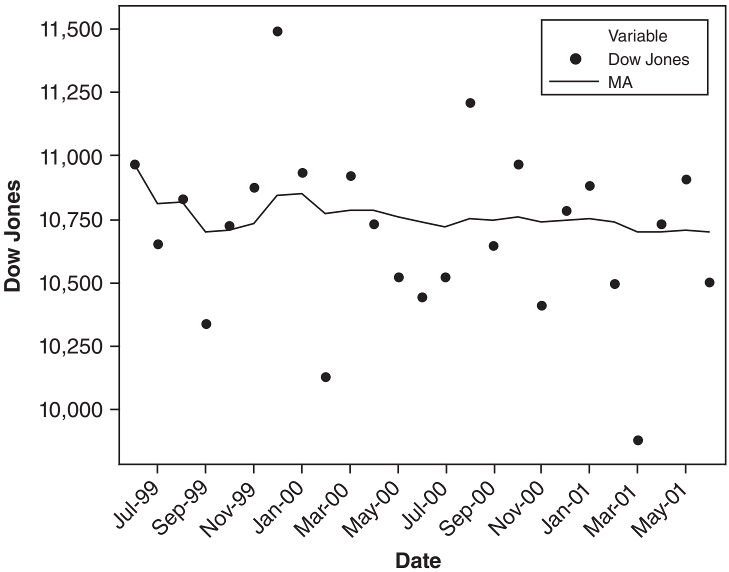


FIGURE 4.2 The Dow Jones Index from June 1999 to June 2001.

The least squares estimate of μ can be found by setting the derivative of SS with respect to μ to 0. This gives

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t, \tag{4.2}$$

where $\hat{\mu}$ is the least squares estimate of μ. Equation (4.2) shows that the least squares estimate of μ is indeed the average of observations up to time T.

Figure 4.2 shows the monthly data for the Dow Jones Index from June 1999 to June 2001. Visual inspection suggests that a constant model can be used to describe the general pattern of the data.[1] To further confirm this claim, we use the smoother described in Eq. (4.2) for each data point by taking the average of the available data up to that point in time. The smoothed observations are shown by the line segments in Figure 4.2. It can be seen that the smoother in Eq. (4.2) indeed extracts the main pattern

[1]Please note that for this data the independent errors assumption in the constant process in Eq. (4.1) may have been violated. Remedies to check and handle such violations will be provided in the following chapters.

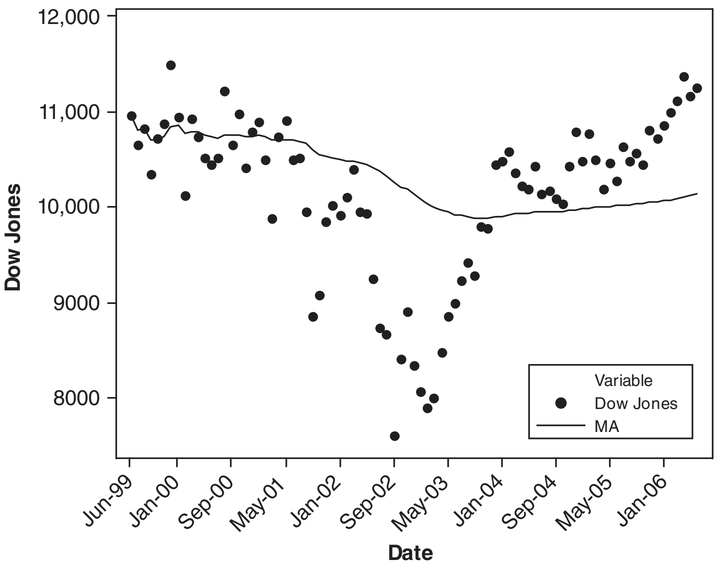


FIGURE 4.3 The Dow Jones Index from June 1999 to June 2006.

in the data and leads to the conclusion that during the 2-year period from June 1999 to June 2001, the Dow Jones Index was quite stable.

As we can see, for the constant process the smoother in Eq. (4.2) is quite effective in providing a clear picture of the underlying pattern. What happens if the process is not constant but exhibits a more complicated pattern? Consider again, for example, the Dow Jones Index from June 1999 to June 2006 given in Figure 4.3 (the complete data set is in Table 4.1). It is clear that the data do not follow the behavior typical of a constant behavior during this period. In Figure 4.3, we can also see the pattern that the smoother in Eq. (4.2) extracts for the same period. As the process changes, this smoother is having trouble keeping up with the process. What could be the reason for the poor performance after June 2001? The answer is quite simple: the constant process assumption is no longer valid. However, as time goes on, the smoother in Eq. (4.2) accumulates more and more data points and gains some sort of “inertia”. So when there is a change in the process, it becomes increasingly more difficult for this smoother to react to it.

How often is the constant process assumption violated? The answer to this question is provided by the Second Law of Thermodynamics, which in the most simplistic way states that if left on its own (free of external influences) any system will deteriorate. Thus the constant process is not

TABLE 4.1 Dow Jones Index at the End of the Month from June 1999 to June 2006

Date	Dow Jones	Date	Dow Jones	Date	Dow Jones	Date	Dow Jones
Jun-99	10,970.8	Apr-01	10,735	Feb-03	7891.08	Dec-04	10,783
Jul-99	10,655.2	May-01	10,911.9	Mar-03	7992.13	Jan-05	10,489.9
Aug-99	10,829.3	Jun-01	10,502.4	Apr-03	8480.09	Feb-05	10,766.2
Sep-99	10,337	Jul-01	10,522.8	May-03	8850.26	Mar-05	10,503.8
Oct-99	10,729.9	Aug-01	9949.75	Jun-03	8985.44	Apr-05	10,192.5
Nov-99	10,877.8	Sep-01	8847.56	Jul-03	9233.8	May-05	10,467.5
Dec-99	11,497.1	Oct-01	9075.14	Aug-03	9415.82	Jun-05	10,275
Jan-00	10,940.5	Nov-01	9851.56	Sep-03	9275.06	Jul-05	10,640.9
Feb-00	10,128.3	Dec-01	10,021.6	Oct-03	9801.12	Aug-05	10,481.6
Mar-00	10,921.9	Jan-02	9920	Nov-03	9782.46	Sep-05	10,568.7
Apr-00	10,733.9	Feb-02	10,106.1	Dec-03	10,453.9	Oct-05	10,440.1
May-00	10,522.3	Mar-02	10,403.9	Jan-04	10488.1	Nov-05	10,805.9
Jun-00	10,447.9	Apr-02	9946.22	Feb-04	10,583.9	Dec-05	10,717.5
Jul-00	10,522	May-02	9925.25	Mar-04	10,357.7	Jan-06	10,864.9
Aug-00	11,215.1	Jun-02	9243.26	Apr-04	10,225.6	Feb-06	10,993.4
Sep-00	10,650.9	Jul-02	8736.59	May-04	10,188.5	Mar-06	11,109.3
Oct-00	10,971.1	Aug-02	8663.5	Jun-04	10,435.5	Apr-06	11,367.1
Nov-00	10,414.5	Sep-02	7591.93	Jul-04	10,139.7	May-06	11,168.3
Dec-00	10,788	Oct-02	8397.03	Aug-04	10,173.9	Jun-06	11,247.9
Jan-01	10,887.4	Nov-02	8896.09	Sep-04	10,080.3		
Feb-01	10,495.3	Dec-02	8341.63	Oct-04	10,027.5		
Mar-01	9878.78	Jan-03	8053.81	Nov-04	10,428		

the norm but at best an exception. So what can we do to deal with this issue? Recall that the problem with the smoother in Eq. (4.2) was that it reacted too slowly to process changes because of its inertia. In fact, when there is a change in the process, earlier data no longer carry the information about the change in the process, yet they contribute to this inertia at an equal proportion compared to the more recent (and probably more useful) data. The most obvious choice is to somehow discount the older data. Also recall that in a simple average, as in Eq. (4.2), all the observations are weighted equally and hence have the same amount of influence on the average. Thus, if the weights of each observation are changed so that earlier observations are weighted less, a faster reacting smoother should be obtained. As mentioned in Section 2.2.2, a common solution is to use the **simple moving average** given in Eq. (2.3):

$$M_T = \frac{y_T + y_{T-1} + \cdots + y_{T-N+1}}{N} = \frac{1}{N} \sum_{t=T-N+1}^{N} y_t.$$

The most crucial issue in simple moving averages is the choice of the **span**, N. A simple moving average will react faster to the changes if N is small. However, we know from Section 2.2.2 that the variance of the simple moving average with uncorrelated observations with variance σ^2 is given as

$$\text{Var}(M_T) = \frac{\sigma^2}{N}.$$

This means that as N gets small, the variance of the moving average gets bigger. This creates a dilemma in the choice of N. If the process is expected to be constant, a large N can be used whereas a small N is preferred if the process is changing. In Figure 4.4, we show the effect of going from a span of 10 observations to 5 observations. While the latter exhibits a more jittery behavior, it nevertheless follows the actual data more closely. A more thorough analysis on the choice of N can be performed based on the prediction error. We will explore this for exponential smoothers in Section 4.6.1, where we will discuss forecasting using exponential smoothing.

A final note on the moving average is that even if the individual observations are independent, the moving averages will be autocorrelated as two successive moving averages contain the same $N-1$ observations. In fact,

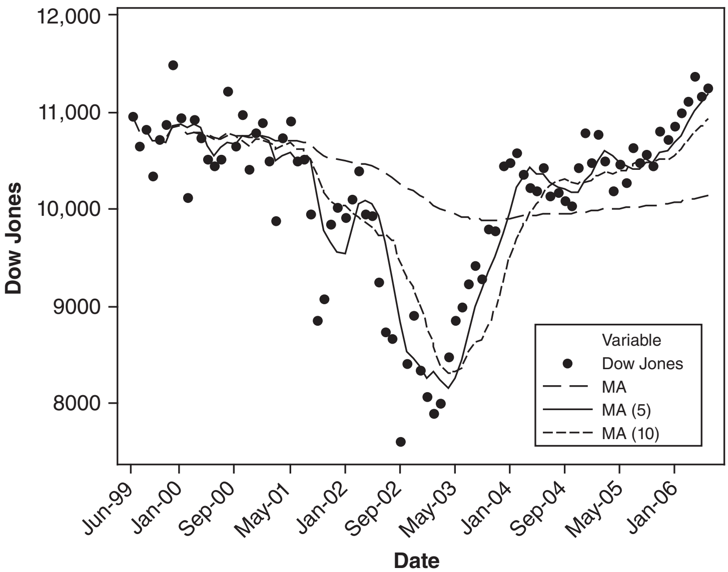


FIGURE 4.4 The Dow Jones Index from June 1999 to June 2006 with moving averages of span 5 and 10.

the autocorrelation function (ACF) of the moving averages that are k-lags apart is given as

$$\rho k = \begin{cases} 1 - \dfrac{|k|}{N}, & k < N \\ 0, & k \geq N \end{cases} .$$

4.2 FIRST-ORDER EXPONENTIAL SMOOTHING

Another approach to obtain a smoother that will react to process changes faster is to give geometrically decreasing weights to the past observations. Hence an exponentially weighted smoother is obtained by introducing a discount factor θ as

$$\sum_{t=0}^{T-1} \theta^t y_{T-t} = y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1. \tag{4.3}$$

Please note that if the past observations are to be discounted in a geometrically decreasing manner, then we should have $|\theta| < 1$. However, the smoother in Eq. (4.3) is not an *average* as the sum of the weights is

$$\sum_{t=0}^{T-1} \theta^t = \frac{1-\theta^T}{1-\theta} \tag{4.4}$$

and hence does not necessarily add up to 1. For that we can adjust the smoother in Eq. (4.3) by multiplying it by $(1-\theta)/(1-\theta^T)$. However, for large T values, θ^T goes to zero and so the exponentially weighted average will have the following form:

$$\begin{aligned} \tilde{y}_T &= (1-\theta)\sum_{t=0}^{T-1} \theta^t y_{T-t} \\ &= (1-\theta)(y_T + \theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1) \end{aligned} \tag{4.5}$$

This is called a **simple** or **first-order exponential smoother**. There is an extensive literature on exponential smoothing. For example, see the books by Brown (1963), Abraham and Ledolter (1983), and Montgomery et al. (1990), and the papers by Brown and Meyer (1961), Chatfield and Yar (1988), Cox (1961), Gardner (1985), Gardner and Dannenbring (1980), and Ledolter and Abraham (1984).

An alternate expression in a recursive form for simple exponential smoothing is given by

$$\begin{aligned} \tilde{y}_T &= (1-\theta)y_T + (1-\theta)(\theta y_{T-1} + \theta^2 y_{T-2} + \cdots + \theta^{T-1} y_1) \\ &= (1-\theta)y_T + \theta \underbrace{(1-\theta)(y_{T-1} + \theta^1 y_{T-2} + \cdots + \theta^{T-2} y_1)}_{\tilde{y}_{T-1}} \\ &= (1-\theta)y_T + \theta \tilde{y}_{T-1}. \end{aligned} \tag{4.6}$$

The recursive form in Eq. (4.6) shows that first-order exponential smoothing can also be seen as the linear combination of the current observation and the smoothed observation at the previous time unit. As the latter contains the data from all previous observations, the smoothed observation at time T is in fact the linear combination of the current observation and the discounted sum of all previous observations. The simple exponential smoother is often represented in a different form by setting $\lambda = 1-\theta$,

$$\tilde{y}_T = \lambda y_T + (1-\lambda)\tilde{y}_{T-1} \tag{4.7}$$

In this representation the **discount factor**, λ, represents the weight put on the last observation and $(1-\lambda)$ represents the weight put on the smoothed value of the previous observations.

Analogous to the size of the span in moving average smoothers, an important issue for the exponential smoothers is the choice of the discount factor, λ. Moreover, from Eq. (4.7), we can see that the calculation of $\tilde{y}_1$ would require us to know $\tilde{y}_0$. We will discuss these issues in the next two sections.

4.2.1 The Initial Value, $\tilde{y}_0$

Since $\tilde{y}_0$ is needed in the recursive calculations that start with $\tilde{y}_1 = \lambda y_1 + (1-\lambda)\tilde{y}_0$, its value needs to be estimated. But from Eq. (4.7) we have

$$
\begin{aligned}
\tilde{y}_1 &= \lambda y_1 + (1-\lambda)\tilde{y}_0 \\
\tilde{y}_2 &= \lambda y_2 + (1-\lambda)\tilde{y}_1 = \lambda y_2 + (1-\lambda)(\lambda y_1 + (1-\lambda)\tilde{y}_0) \\
&= \lambda(y_2 + (1-\lambda)y_1) + (1-\lambda)^2\tilde{y}_0 \\
\tilde{y}_3 &= \lambda(y_3 + (1-\lambda)y_2 + (1-\lambda)^2 y_1) + (1-\lambda)^3\tilde{y}_0 \\
&\vdots \\
\tilde{y}_T &= \lambda(y_T + (1-\lambda)y_{T-1} + \cdots + (1-\lambda)^{T-1}y_1) + (1-\lambda)^T\tilde{y}_0,
\end{aligned}
$$

which means that as T gets large and hence $(1-\lambda)^T$ gets small, the contribution of $\tilde{y}_0$ to $\tilde{y}_T$ becomes negligible. Thus for large data sets, the estimation of $\tilde{y}_0$ has little relevance. Nevertheless, two commonly used estimates for $\tilde{y}_0$ are the following.

1. Set $\tilde{y}_0 = y_1$. If the changes in the process are expected to occur early and fast, this choice for the starting value for $\tilde{y}_T$ is reasonable.
2. Take the average of the available data or a subset of the available data, $\bar{y}$, and set $\tilde{y}_0 = \bar{y}$. If the process is at least at the beginning locally constant, this starting value may be preferred.

4.2.2 The Value of λ

In Figures 4.5 and 4.6, respectively, we have two simple exponential smoothers for the Dow Jones Index data with $\lambda = 0.2$ and $\lambda = 0.4$. It can be seen that in the latter the smoothed values follow the original observations more closely. In general, as λ gets closer to 1, and more emphasis is put on the last observation, the smoothed values will approach the original observations. Two extreme cases will be when $\lambda = 0$ and $\lambda = 1$. In the former, the smoothed values will all be equal to a constant, namely, y_0.

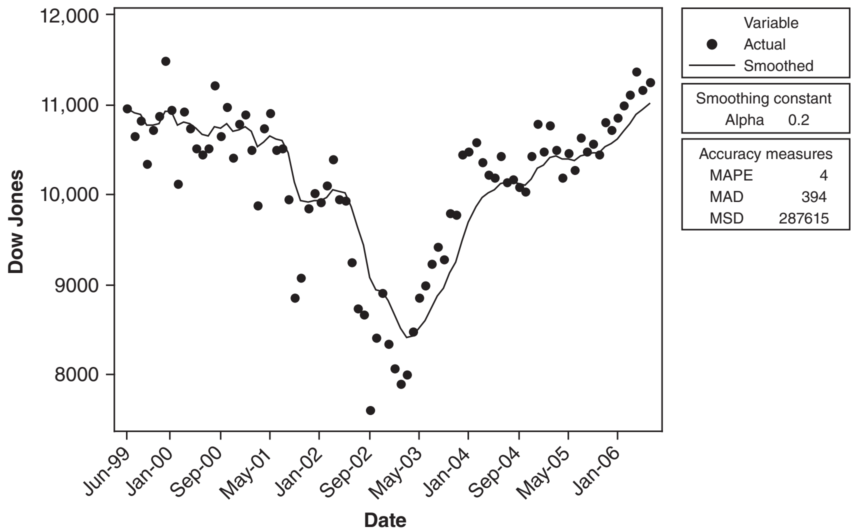


FIGURE 4.5 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$.

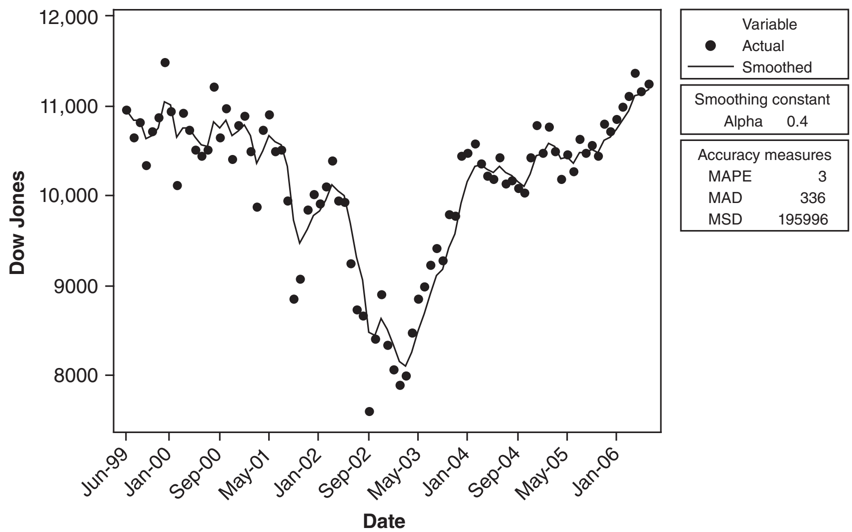


FIGURE 4.6 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$.

We can think of the constant line as the "smoothest" version of whatever pattern the actual time series follows. For $\lambda = 1$, we have $\tilde{y}_T = y_T$ and this will represent the "least" smoothed (or unsmoothed) version of the original time series. We can accordingly expect the variance of the simple exponential smoother to vary between 0 and the variance of the original time series based on the choice of λ. Note that under the independence and constant variance assumptions we have

$$\begin{aligned}
\mathrm{Var}(\tilde{y}_T) &= \mathrm{Var}\left(\lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t}\right) \\
&= \lambda^2 \sum_{t=0}^{\infty} (1-\lambda)^{2t} \mathrm{Var}\,(y_{T-t}) \\
&= \lambda^2 \sum_{t=0}^{\infty} (1-\lambda)^{2t} \mathrm{Var}(y_T) \\
&= \mathrm{Var}(y_T)\lambda^2 \sum_{t=0}^{\infty} (1-\lambda)^{2t} \\
&= \frac{\lambda}{(2-\lambda)} \mathrm{Var}(y_T).
\end{aligned} \tag{4.8}$$

Thus the question will be how much smoothing is needed. In the literature, λ values between 0.1 and 0.4 are often recommended and do indeed perform well in practice. A more rigorous method of finding the right λ value will be discussed in Section 4.6.1.

Example 4.1 Consider the Dow Jones Index from June 1999 to June 2006 given in Figure 4.3. For first-order exponential smoothing we would need to address two issues as stated in the previous sections: how to pick the initial value y_0 and the smoothing constant λ. Following the recommendation in Section 4.2.2, we will consider the smoothing constants 0.2 and 0.4. As for the initial value, we will consider the first recommendation in Section 4.2.1 and set $\tilde{y}_0 = y_1$. Figures 4.5 and 4.6 show the smoothed and actual data obtained from Minitab with smoothing constants 0.2 and 0.4, respectively.

Note that Minitab reports several measures of accuracy; MAPE, MAD, and MSD. Mean absolute percentage error (MAPE) is the average absolute percentage change between the predicted value that is $\tilde{y}_{t-1}$ for a one-step-ahead forecast and the true value, given as

$$\mathrm{MAPE} = \frac{\sum_{t=1}^{T} |(y_t - \tilde{y}_{t-1})/y_t|}{T} \times 100 \quad (y_t \neq 0).$$

Mean absolute deviation (MAD) is the average absolute difference between the predicted and the true values, given as

$$\mathrm{MAD} = \frac{\sum_{t=1}^{T} |(y_t - \tilde{y}_{t-1})|}{T}.$$

Mean squared deviation (MSD) is the average squared difference between the predicted and the true values, given as

$$\mathrm{MSD} = \frac{\sum_{t=1}^{T} (y_t - \tilde{y}_{t-1})^2}{T}.$$

It should also be noted that the smoothed data with $\lambda = 0.4$ follows the actual data closer. However, in both cases, when there is an apparent linear trend in the data (e.g., from February 2003 to February 2004) the smoothed values consistently underestimate the actual data. We will discuss this issue in greater detail in Section 4.3.

As an alternative estimate for the initial value, we can also use the average of the data between June 1999 and June 2001, since during this period the time series data appear to be stable. Figures 4.7 and 4.8 show

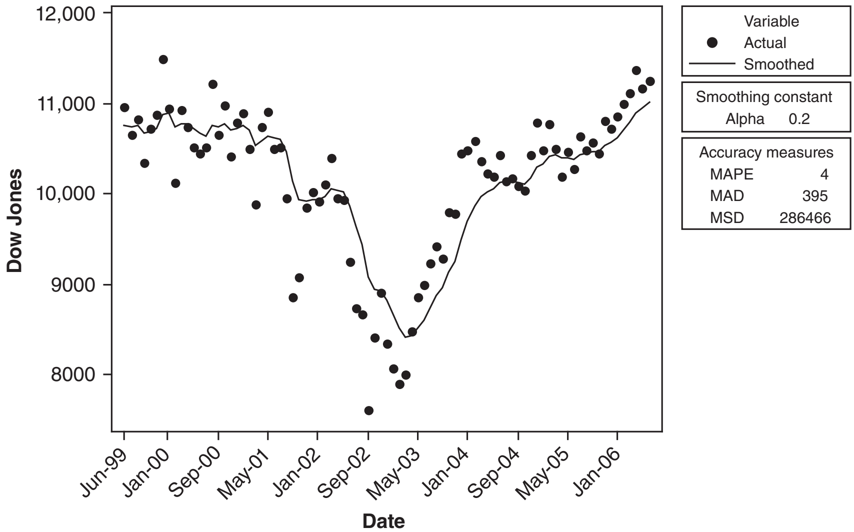


FIGURE 4.7 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.2$ and $\tilde{y}_0 = (\sum_{t=1}^{25} y_t/25)$ (i.e., initial value equal to the average of the first 25 observations).

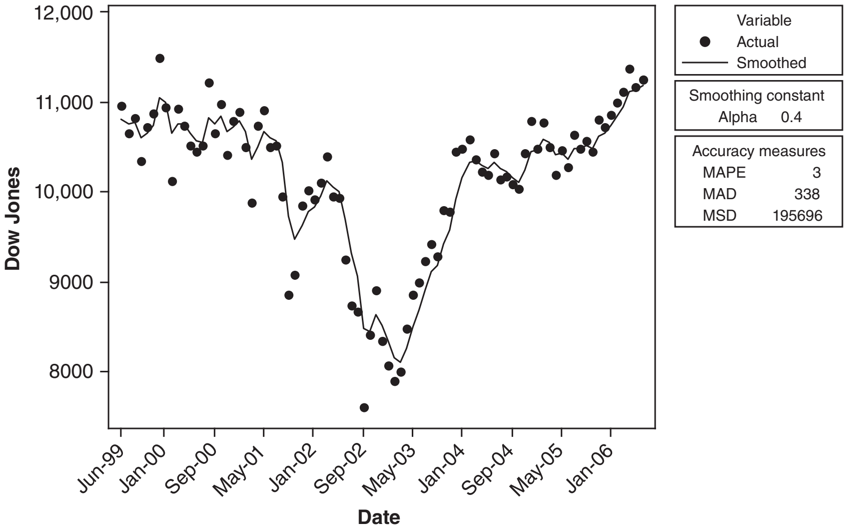


FIGURE 4.8 The Dow Jones Index from June 1999 to June 2006 with first-order exponential smoothing with $\lambda = 0.4$ and $\tilde{y}_0 = (\sum_{t=1}^{25} y_t/25)$ (i.e., initial value equal to the average of the first 25 observations).

the single exponential smoothing with the initial value equal to the average of the first 25 observations corresponding to the period between June 1999 and June 2001. Note that the choice of the initial value has very little effect on the smoothed values as time goes on.

4.3 MODELING TIME SERIES DATA

In Section 4.1, we considered the constant process where the time series data are expected to vary around a constant level with random fluctuations, which are usually characterized by uncorrelated errors with mean 0 and constant variance σ_ε^2. In fact the constant process represents a very special case in a more general set of models often used in modeling time series data as a function of time. The general class of models can be represented as

$$y_t = f(t; \beta) + \varepsilon_\mathrm{t}, \tag{4.9}$$

where β is the vector of unknown parameters and ε_t represents the uncorrelated errors. Thus as a member of this general class of models, the constant process can be represented as

$$y_t = \beta_0 + \varepsilon_\mathrm{t}, \tag{4.10}$$

where β_0 is equal to μ in Eq. (4.1). We have seen in Chapter 3 how to estimate and make inferences about the regression coefficients. The same principles apply to the class of models in Eq. (4.9). However, we have seen in Section 4.1 that the least squares estimates for β_0 at any given time T will be very slow to react to changes in the level of the process. For that, we suggested to use either the moving average or simple exponential smoothing.

As mentioned earlier, smoothing techniques are effective in illustrating the underlying pattern in the time series data. We have so far focused particularly on exponential smoothing techniques. For the class of models given in Eq. (4.9), we can find another use for the exponential smoothers: model estimation. Indeed for the constant process, we can see the simple exponential smoother as the estimate of the process level, or in regards to Eq. (4.10) an estimate of β_0. To show this in greater detail we need to introduce the sum of weighted squared errors for the constant process. Remember that the sum of squared errors for the constant process is given by

$$SS_E = \sum_{t=1}^{T} (y_t - \mu)^2.$$

If we argue that not all observations should have equal influence on the sum and decide to introduce a string of weights that are geometrically decreasing in time, the sum of squared errors becomes

$$SS_E^* = \sum_{t=0}^{T-1} \theta^t (y_{T-1} - \beta_0)^2, \tag{4.11}$$

where $|\theta|\ 1 < 1$. To find the least squares estimate for β_0, we take the derivative of Eq. (4.11) with respect to β_0 and set it to zero:

$$\left.\frac{dSS_E^*}{d\beta_0}\right|_{\beta_0} = -2\sum_{t=0}^{T-1} \theta^t (y_{T-t} - \hat{\beta}_0) = 0. \tag{4.12}$$

The solution to Eq. (4.12), $\hat{\beta}_0$, which is the least squares estimate of β_0, is

$$\hat{\beta}_0 \sum_{t=0}^{T-1} \theta^t = \sum_{t=0}^{T-1} \theta^t y_{T-t}. \tag{4.13}$$

From Eq. (4.4), we have

$$\hat{\beta}_0 = \frac{1-\theta}{1-\theta^T} \sum_{t=0}^{T-1} \theta^t y_{T-t}. \tag{4.14}$$

Once again for large T, θ^T goes to zero. We then have

$$\hat{\beta}_0 = (1-\theta) \sum_{t=0}^{T-1} \theta^t y_{T-t}. \tag{4.15}$$

We can see from Eqs. (4.5) and (4.15) that $\beta_0 = \tilde{y}_T$. Thus the simple exponential smoothing procedure does in fact provide a weighted least squares estimate of β_0 in the constant process with weights that are exponentially decreasing in time.

Now we return to our general class of models given in Eq. (4.9) and note that $f(t; \beta)$ can in fact be any function of t. For practical purposes it is usually more convenient to consider the polynomial family for nonseasonal time series. For seasonal time series, we will consider other forms of $f(t; \beta)$ that fit the data and exhibit a certain periodicity better. In the polynomial family, the constant process is indeed the simplest model we can consider. We will now consider the next obvious choice: the linear trend model.

4.4 SECOND-ORDER EXPONENTIAL SMOOTHING

We will now return to our Dow Jones Index data but consider only the subset of the data from February 2003 to February 2004 as given in Figure 4.9. Evidently for that particular time period it was a bullish market and correspondingly the Dow Jones Index exhibits an upward linear trend as indicated with the dashed line.

For this time period, an appropriate model in time from the polynomial family should be the linear trend model given as

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{4.16}$$

where the ε_t is once again assumed to be uncorrelated with mean 0 and constant variance $\sigma^2{}_\varepsilon$. Based on what we have learned so far, we may attempt to smooth/model this linear trend using the simple exponential smoothing procedure. The actual and fitted values for the simple exponential smoothing procedure are given in Figure 4.10. For the exponential

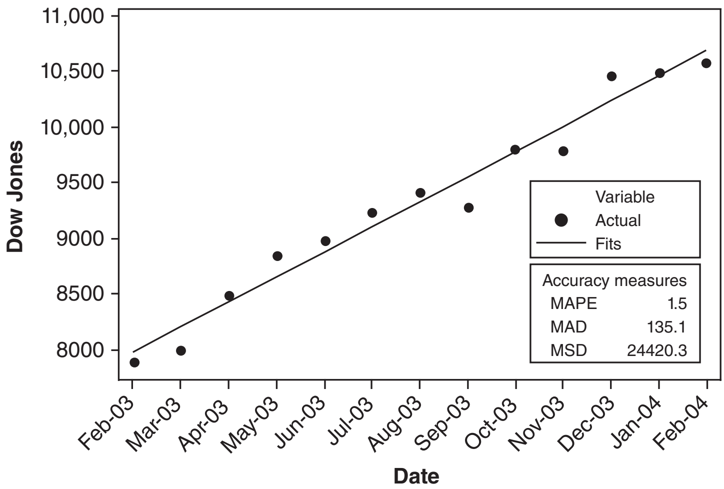


FIGURE 4.9 The Dow Jones Index from February 2003 to February 2004.

smoother, without any loss of generality, we used $\tilde{y}_0 = y_1$ and $\lambda = 0.3$. From Figure 4.10, we can see that while the simple exponential smoother was to some extent able to capture the slope of the linear trend, it also exhibits some bias. That is, the fitted values based on the exponential smoother are consistently underestimating the actual data. More interestingly, the amount of underestimation is more or less constant for all observations.

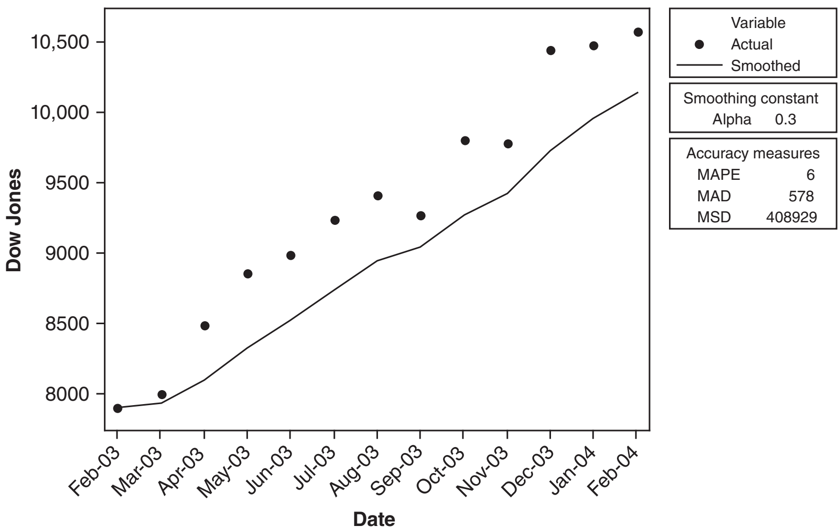


FIGURE 4.10 The Dow Jones Index from February 2003 to February 2004 with simple exponential smoothing with $\lambda = 0.3$.

In fact similar behavior for the simple exponential smoother can be observed in Figure 4.5 for the entire data from June 1999 to June 2006. Whenever the data exhibit a linear trend, the simple exponential smoother seems to over- or underestimates the actual data consistently. To further explore this, we will consider the expected value of $\tilde{y}_T$,

$$\begin{aligned} E(\tilde{y}_T) &= E\left(\lambda \sum_{t=0}^{\infty} (1-\lambda)^t y_{T-t}\right) \\ &= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t E(y_{T-t}). \end{aligned}$$

For the linear trend model in Eq. (4.16), $E\,(y_t) = \beta_0 + \beta_1 t$. So we have

$$\begin{aligned} E(\tilde{y}_T) &= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_0 + \beta_1 (T-t)) \\ &= \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_0 + \beta_1 T) - \lambda \sum_{t=0}^{\infty} (1-\lambda)^t (\beta_1 t) \\ &= (\beta_0 + \beta_1 T)\lambda \sum_{t=0}^{\infty} (1-\lambda)^t - \lambda \beta_1 \sum_{t=0}^{\infty} (1-\lambda)^t t. \end{aligned}$$

But for the infinite sums we have

$$\sum_{t=0}^{\infty} (1-\lambda)^t = \frac{1}{1-(1-\lambda)} = \frac{1}{\lambda} \text{ and } \sum_{t=0}^{\infty} (1-\lambda)^t t = \frac{1-\lambda}{\lambda^2}.$$

Hence the expected value of the simple exponential smoother for the linear trend model is

$$\begin{aligned} E(\tilde{y}_T) &= (\beta_0 + \beta_1 T) - \frac{1-\lambda}{\lambda}\beta_1 \\ &= E(y_T) - \frac{1-\lambda}{\lambda}\beta_1. \end{aligned} \tag{4.17}$$

This means that the simple exponential smoother is a biased estimator for the linear trend model and the amount of bias is $-[(1-\lambda)/\lambda]\beta_1$. This indeed explains the underestimation in Figure 4.10. One solution will be to use a large λ value since $(1-\lambda)/\lambda \to 0$ as $\lambda \to 1$. In Figure 4.11, we show two simple exponential smoothers with $\lambda = 0.3$ and $\lambda = 0.99$. It can be

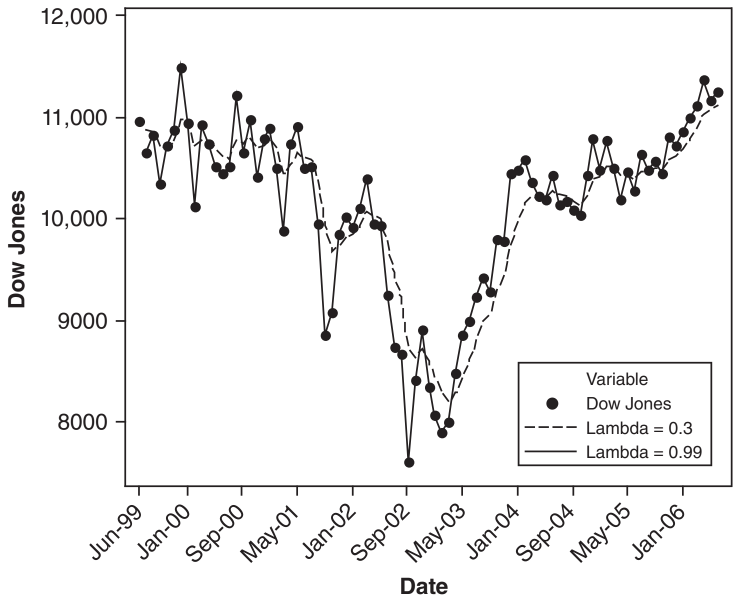


FIGURE 4.11 The Dow Jones Index from June 1999 to June 2006 using exponential smoothing with $\lambda = 0.3$ and 0.99.

seen that the latter does a better job in capturing the linear trend. However, it should also be noted that as the smoother with $\lambda = 0.99$ follows the actual observations very closely, it fails to smooth out the constant pattern during the first 2 years of the data. A method based on adaptive updating of the discount factor, λ, following the changes in the process is given in Section 4.6.4. In this section to model a linear trend model we will instead introduce the second-order exponential smoothing by applying simple exponential smoothing on $\tilde{y}_T$ as

$$\tilde{y}_T^{(2)} = \lambda \tilde{y}_T^{(1)} + (1 - \lambda)\tilde{y}_{T-1}^{(2)}, \tag{4.18}$$

where $\tilde{y}_T^{(1)}$ and $\tilde{y}_T^{(2)}$ denote the first- and second-order smoothed exponentials, respectively. Of course, in Eq. (4.18) we can use a different λ than in Eq. (4.7). However, for the derivations that follow, we will assume that the same λ is used in the calculations of both $\tilde{y}_T^{(1)}$ and $\tilde{y}_T^{(2)}$.

From Eq. (4.17), we can see that the first-order exponential smoother introduces bias in estimating a linear trend. It can also be seen in Figure 4.7 that the first-order exponential smoother for the linear trend model exhibits a linear trend as well. Hence the second-order smoother—that is, a first-order exponential smoother of the original first-order exponential smoother—should also have a bias. We can represent this as

$$E\left(\tilde{y}_T^{(2)}\right) = E\left(\tilde{y}_T^{(1)}\right) - \frac{1-\lambda}{\lambda}\beta_1. \tag{4.19}$$

From Eq. (4.19), an estimate for β_1 at time T is

$$\hat{\beta}_{1,T} = \frac{\lambda}{1-\lambda}\left(\tilde{y}_T^1 - \tilde{y}_T^2\right) \tag{4.20}$$

and for an estimate of β_0 at time T, we have from Eq. (4.17)

$$\begin{aligned} \tilde{y}_T^{(1)} &= \left(\hat{\beta}_{0,T} + \hat{\beta}_{1,T}T\right) - \frac{1-\lambda}{\lambda}\hat{\beta}_{1,T} \\ \Rightarrow \hat{\beta}_{0,T} &= \tilde{y}_T^{(1)} - T\hat{\beta}_{1,T} + \frac{1-\lambda}{\lambda}\hat{\beta}_{1,T}. \end{aligned} \tag{4.21}$$

In terms of the first- and second-order exponential smoothers, we have

$$\begin{aligned} \hat{\beta}_{0,T} &= \tilde{y}_T^{(1)} - T\frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) + \frac{1-\lambda}{\lambda}\left(\frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right)\right) \\ &= \tilde{y}_T^{(1)} - T\frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) + \left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) \\ &= \left(2 - T\frac{\lambda}{1-\lambda}\right)\tilde{y}_T^{(1)} - \left(1 - T\frac{\lambda}{1-\lambda}\right)\tilde{y}_T^{(2)}. \end{aligned} \tag{4.22}$$

Finally, combining Eq. (4.20) and (4.22), we have a predictor for y_T as

$$\begin{aligned} \tilde{y}_T &= \hat{\beta}_{0,T} + \hat{\beta}_{1,T}T \\ &= 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}. \end{aligned} \tag{4.23}$$

It can easily be shown that $\hat{y}_T$ is an unbiased predictor of y_T. In Figure 4.12, we use Eq. (4.23) to estimate the Dow Jones Index from February 2003 to February 2004. From Figures 4.10 and 4.12, we can clearly see that the second-order exponential smoother is doing a much better job in modeling the linear trend compared to the simple exponential smoother.

As in the simple exponential smoothing, we have the same two issues to deal with: initial values for the smoothers and the discount factors. The

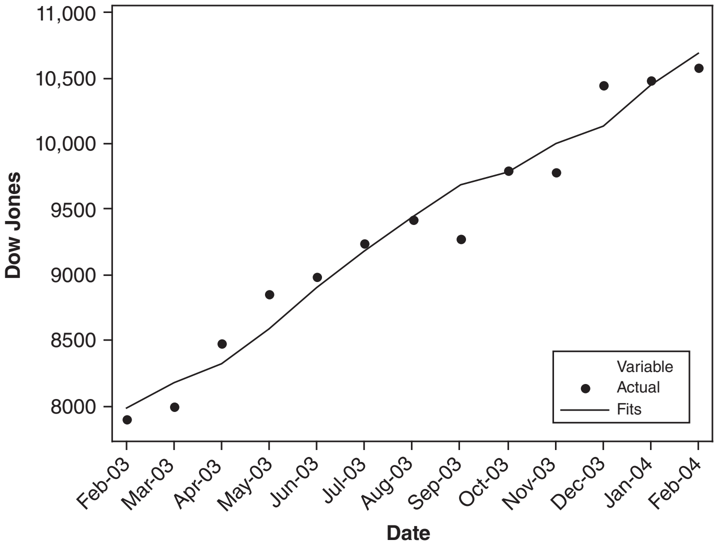


FIGURE 4.12 The Dow Jones Index from February 2003 to February 2004 with second-order exponential smoother with discount factor of 0.3.

latter will be discussed in Section 4.6.1. For the former we will combine Eqs. (4.17) and (4.19) as the following:

$$\begin{aligned}\tilde{y}_0^{(1)} &= \hat{\beta}_{0,0} - \frac{1-\lambda}{\lambda}\hat{\beta}_{1,0} \\ \tilde{y}_0^{(2)} &= \hat{\beta}_{0,0} - 2\left(\frac{1-\lambda}{\lambda}\right)\hat{\beta}_{1,0}.\end{aligned} \tag{4.24}$$

The initial estimates of the model parameters are usually obtained by fitting the linear trend model to the entire or a subset of the available data. The least squares estimates of the parameter estimates are then used for $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$.

Example 4.2 Consider the US Consumer Price Index (CPI) from January 1995 to December 2004 in Table 4.2. Figure 4.13 clearly shows that the data exhibits a linear trend. To smooth the data, following the recommendation in Section 4.2, we can use single exponential smoothing with $\lambda = 0.3$ as given in Figure 4.14.

As we expected, the exponential smoother does a very good job in capturing the general trend in the data and provides a less jittery (smooth) version of it. However, we also notice that the smoothed values are

TABLE 4.2 Consumer Price Index from January 1995 to December 2004

Month-Year	CPI	Month-Year	CPI	Month-Year	CPI	Month-Year	CPI	Month-Year	CPI
Jan-1995	150.3	Jan-1997	159.1	Jan-1999	164.3	Jan-2001	175.1	Jan-2003	181.7
Feb-1995	150.9	Feb-1997	159.6	Feb-1999	164.5	Feb-2001	175.8	Feb-2003	183.1
Mar-1995	151.4	Mar-1997	160	Mar-1999	165	Mar-2001	176.2	Mar-2003	184.2
Apr-1995	151.9	Apr-1997	160.2	Apr-1999	166.2	Apr-2001	176.9	Apr-2003	183.8
May-1995	152.2	May-1997	160.1	May-1999	166.2	May-2001	177.7	May-2003	183.5
Jun-1995	152.5	Jun-1997	160.3	Jun-1999	166.2	Jun-2001	178	Jun-2003	183.7
Jul-1995	152.5	Jul-1997	160.5	Jul-1999	166.7	Jul-2001	177.5	Jul-2003	183.9
Aug-1995	152.9	Aug-1997	160.8	Aug-1999	167.1	Aug-2001	177.5	Aug-2003	184.6
Sep-1995	153.2	Sep-1997	161.2	Sep-1999	167.9	Sep-2001	178.3	Sep-2003	185.2
Oct-1995	153.7	Oct-1997	161.6	Oct-1999	168.2	Oct-2001	177.7	Oct-2003	185
Nov-1995	153.6	Nov-1997	161.5	Nov-1999	168.3	Nov-2001	177.4	Nov-2003	184.5
Dec-1995	153.5	Dec-1997	161.3	Dec-1999	168.3	Dec-2001	176.7	Dec-2003	184.3
Jan-1996	154.4	Jan-1998	161.6	Jan-2000	168.8	Jan-2002	177.1	Jan-2004	185.2
Feb-1996	154.9	Feb-1998	161.9	Feb-2000	169.8	Feb-2002	177.8	Feb-2004	186.2
Mar-1996	155.7	Mar-1998	162.2	Mar-2000	171.2	Mar-2002	178.8	Mar-2004	187.4
Apr-1996	156.3	Apr-1998	162.5	Apr-2000	171.3	Apr-2002	179.8	Apr-2004	188
May-1996	156.6	May-1998	162.8	May-2000	171.5	May-2002	179.8	May-2004	189.1
Jun-1996	156.7	Jun-1998	163	Jun-2000	172.4	Jun-2002	179.9	Jun-2004	189.7
Jul-1996	157	Jul-1998	163.2	Jul-2000	172.8	Jul-2002	180.1	Jul-2004	189.4
Aug-1996	157.3	Aug-1998	163.4	Aug-2000	172.8	Aug-2002	180.7	Aug-2004	189.5
Sep-1996	157.8	Sep-1998	163.6	Sep-2000	173.7	Sep-2002	181	Sep-2004	189.9
Oct-1996	158.3	Oct-1998	164	Oct-2000	174	Oct-2002	181.3	Oct-2004	190.9
Nov-1996	158.6	Nov-1998	164	Nov-2000	174.1	Nov-2002	181.3	Nov-2004	191
Dec-1996	158.6	Dec-1998	163.9	Dec-2000	174	Dec-2002	180.9	Dec-2004	190.3

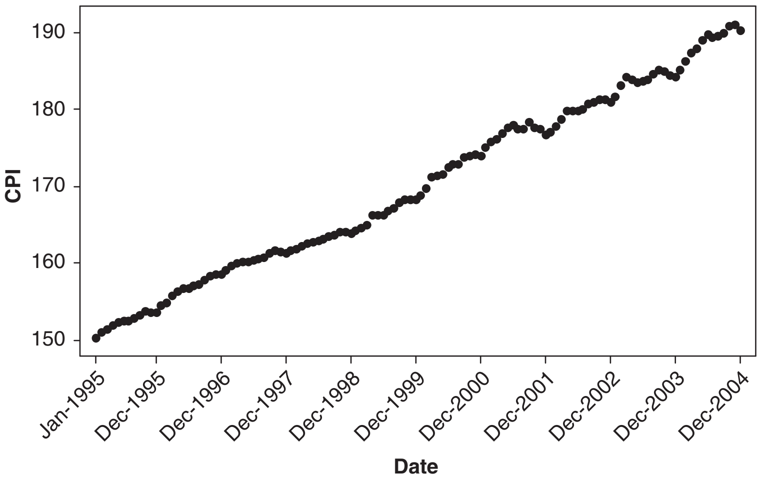


FIGURE 4.13 US Consumer Price Index from January 1995 to December 2004.

consistently below the actual values. Hence there is an apparent bias in our smoothing. To fix this problem we have two choices: use a bigger λ or **second-order** exponential smoothing. The former will lead to less smooth estimates and hence defeat the purpose. For the latter, however, we can use $\lambda = 0.3$ to calculate and $\tilde{y}_T^{(1)}$ and $\tilde{y}_T^{(2)}$ as given in Table 4.3.

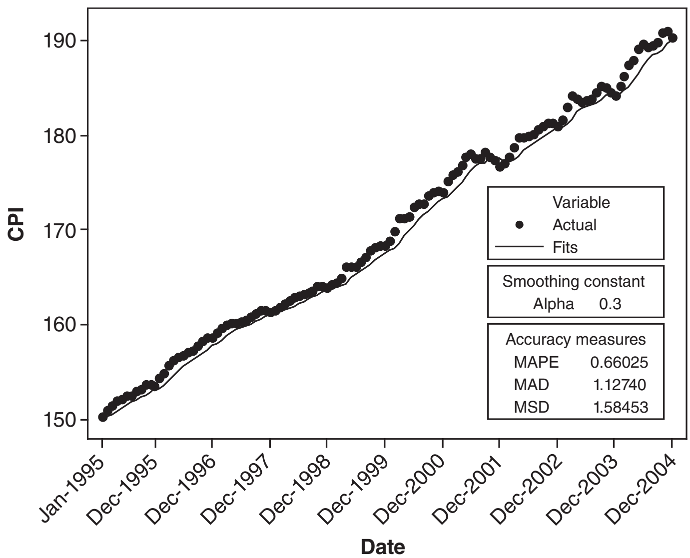


FIGURE 4.14 Single exponential smoothing of the US Consumer Price Index (with $\tilde{y}_0 = y_1$).

TABLE 4.3 Second-Order Exponential Smoothing of the US Consumer Price Index (with $\lambda = 0.3, \tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_0^{(1)}$

Date	y_t	$\tilde{y}_T^{(1)}$	$\tilde{y}_T^{(2)}$	$\tilde{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$
Jan-1995	150.3	150.300	150.300	150.300
Feb-1995	150.9	150.480	150.354	150.606
Mar-1995	151.4	150.756	150.475	151.037
Apr-1995	151.9	151.099	150.662	151.536
May-1995	152.2	151.429	150.892	151.967
Nov-2004	191.0	190.041	188.976	191.106
Dec-2004	190.3	190.119	189.319	190.919

Note that we used $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_0^{(1)}$ as the initial values of $\tilde{y}_T^{(1)}$ and $\tilde{y}_T^{(2)}$. A more rigorous approach would involve fitting a linear regression model in time to the available data that give

$$\begin{aligned}\hat{y}_t &= \hat{\beta}_{0,T} + \hat{\beta}_{1,T} t \\ &= 149.89 + 0.33t,\end{aligned}$$

where t goes from 1 to 120. Then from Eq. (4.24) we have

$$\begin{aligned}\tilde{y}_0^{(1)} &= \hat{\beta}_{0,0} - \frac{1-\lambda}{\lambda}\hat{\beta}_{1,0} \\ &= 149.89 - \frac{1-0.3}{0.3}0.33 = 146.22 \\ \tilde{y}_0^{(2)} &= \hat{\beta}_{0,0} - 2\left(\frac{1-\lambda}{\lambda}\right)\hat{\beta}_{1,0} \\ &= 149.89 - 2\left(\frac{1-0.3}{0.3}\right)0.33 = 142.56.\end{aligned}$$

Figure 4.15 shows the second-order exponential smoothing of the CPI. As we can see, the second-order exponential smoothing not only captures the trend in the data but also does not exhibit any bias.

The calculations for the second-order smoothing for the CPI data are performed using Minitab. We first obtained the first-order exponential smoother for the CPI, $\tilde{y}_T^{(1)}$, using $\lambda = 0.3$ and $\tilde{y}_0^{(1)} = y_1$. Then we obtained $\tilde{y}_T^{(2)}$ by taking the first-order exponential smoother $\tilde{y}_T^{(1)}$ using $\lambda = 0.3$ and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$. Then using Eq. (4.23) we have $\hat{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$.

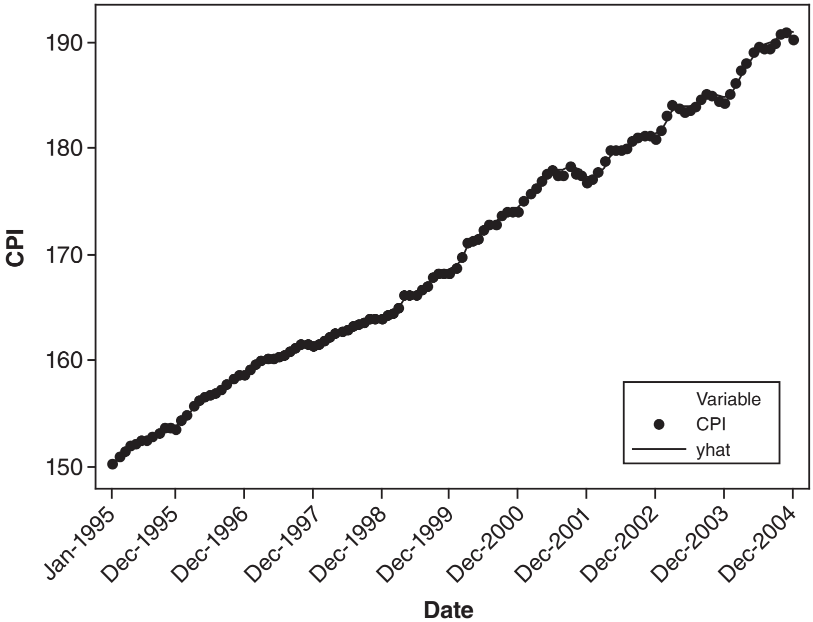


FIGURE 4.15 Second-order exponential smoothing of the US Consumer Price Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

The "Double Exponential Smoothing" option available in Minitab is a slightly different approach based on Holt's method (Holt, 1957). This method divides the time series data into two components: the level, L_t, and the trend, T_t. These two components can be calculated from

$$L_t = \alpha y_t + (1 - \alpha)(L_{t-1} + T_{t-1})$$
$$T_t = \gamma(L_t - L_{t-1}) + (1 - \gamma)T_{t-1}$$

Hence for a given set of α and γ, these two components are calculated and L_t is used to obtain the double exponential smoothing of the data at time t. Furthermore, the sum of the level and trend components at time t can be used as the one-step-ahead $(t + 1)$ forecast. Figure 4.16 shows the actual and smoothed data using the double exponential smoothing option in Minitab with $\alpha = 0.3$ and $\gamma = 0.3$.

In general, the initial values for the level and the trend terms can be obtained by fitting a linear regression model to the CPI data with time as

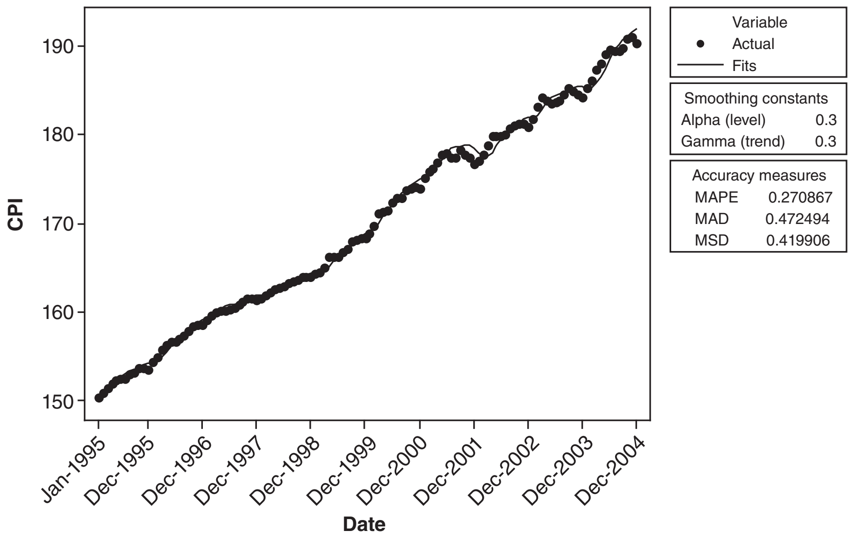


FIGURE 4.16 The double exponential smoothing of the US Consumer Price Index (with $\alpha = 0.3$ and $\gamma = 0.3$).

the regressor. Then the intercept and the slope can be used as the initial values of L_t and T_t respectively.

Example 4.3 For the Dow Jones Index data, we observed that first-order exponential smoothing with low values of λ showed some bias when there were linear trends in the data. We may therefore decide to use the second-order exponential smoothing approach for this data as shown in Figure 4.17. Note that the bias present with first-order exponential smoothing has been eliminated. The calculations for second-order exponential smoothing for the Dow Jones Index are given in Table 4.4.

4.5 HIGHER-ORDER EXPONENTIAL SMOOTHING

So far we have discussed the use of exponential smoothers in estimating the constant and linear trend models. For the former we employed the **simple** or **first-order** exponential smoother and for the latter the **second-order** exponential smoother. It can further be shown that for the general nth-degree polynomial model of the form

$$y_t = \beta_0 + \beta_1 t + \frac{\beta_2}{2!}t^2 + \cdots + \frac{\beta_n}{n!}t^n + \varepsilon_t, \tag{4.25}$$

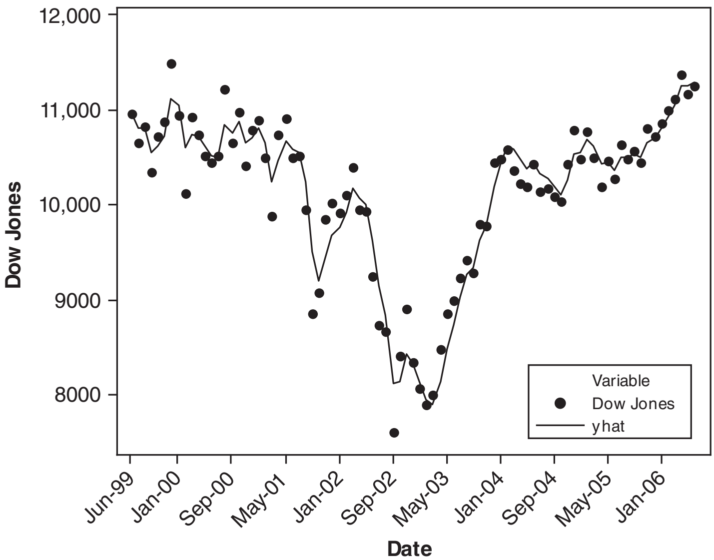


FIGURE 4.17 The second-order exponential smoothing of the Dow Jones Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$).

where the ε_t is assumed to be independent with mean 0 and constant variance $\sigma_\varepsilon{}^2$, we employ $(n+1)$-order exponential smoothers

$$\begin{aligned}
\tilde{y}_T^{(2)} &= \lambda y_T + (1-\lambda)\tilde{y}_{T-1}^{(1)} \\
\tilde{y}_T^{(2)} &= \lambda \tilde{y}_T^{(1)} + (1-\lambda)\tilde{y}_{T-1}^{(2)} \\
&\vdots \\
\tilde{y}_T^{(n)} &= \lambda \tilde{y}_T^{(n-1)} + (1-\lambda)\tilde{y}_{T-1}^{(n)}
\end{aligned}$$

TABLE 4.4 Second-Order Exponential Smoothing of the Dow Jones Index (with $\lambda = 0.3$, $\tilde{y}_0^{(1)} = y_1$, and $\tilde{y}_0^{(2)} = \tilde{y}_1^{(1)}$)

Date	$\tilde{y}_t$	$\tilde{y}_T^1$	$\tilde{y}_T^2$	$\hat{y}_T = 2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}$
Jun-1999	10,970.8	10,970.8	10,970.8	10,970.8
Jul-1999	10,655.2	10,876.1	10,942.4	10,809.8
Aug-1999	10,829.3	10,862.1	10,918.3	10,805.8
Sep-1999	10,337.0	10,704.6	10,854.2	10,554.9
Oct-1999	10,729.9	10,712.2	10,811.6	10,612.7
May-2006	11,168.3	11,069.4	10,886.5	11,252.3
Jun-2006	11,247.9	11,123.0	10,957.4	11,288.5

to estimate the model parameters. For even the quadratic model (second-degree polynomial), the calculations get quite complicated. Refer to Montgomery et al. (1990), Brown (1963), and Abraham and Ledolter (1983) for the solutions to higher-order exponential smoothing problems. If a high-order polynomial does seem to be required for the time series, the autoregressive integrated moving average (ARIMA) models and techniques discussed in Chapter 5 can instead be considered.

4.6 FORECASTING

We have so far considered exponential smoothing techniques as either visual aids to point out the underlying patterns in the time series data or to estimate the model parameters for the class of models given in Eq. (4.9). The latter brings up yet another use of exponential smoothing—forecasting future observations. At time T, we may wish to forecast the observation in the next time unit, $T + 1$, or further into the future. For that, we will denote the τ-step-ahead forecast made at time T as $\hat{y}_{T+\tau}(T)$. In the next two sections and without any loss of generality, we will once again consider first- and second-order exponential smoothers as examples for forecasting time series data from the constant and linear trend processes.

4.6.1 Constant Process

In Section 4.2 we discussed first-order exponential smoothing for the constant process in Eq. (4.1) as

$$\tilde{y}_T = \lambda y_T + (1 - \lambda)\tilde{y}_{T-1}.$$

In Section 4.3 we further showed that the constant level in Eq. (4.1), β_0, can be estimated by $\tilde{y}_T$. Since the constant model consists of two parts—β_0 that can be estimated by the first-order exponential smoother and the random error that cannot be predicted—our forecast for the future observation is simply equal to the current value of the exponential smoother

$$\hat{y}_{T+\tau}(T) = \tilde{y}_T = \tilde{y}_T. \tag{4.26}$$

Please note that, for the constant process, the forecast in Eq. (4.26) is the same for all future values. Since there may be changes in the level of the constant process, forecasting all future observations with the same value will most likely be misleading. However, as we start accumulating more observations, we can update our forecast. For example, if the data at $T + 1$ become available, our forecast for the future observations becomes

$$\tilde{y}_{T+1} = \lambda y_{T+1} + (1 - \lambda)\tilde{y}_T$$

or

$$\hat{y}_{T+1+\tau}(T + 1) = \lambda y_{T+1} + (1 - \lambda)\hat{y}_{T+\tau}(T) \tag{4.27}$$

We can rewrite Eq. (4.27) for $\tau = 1$ as

$$\begin{aligned}\hat{y}_{T+2}(T + 1) &= \hat{y}_{T+1}(T) + \lambda(y_{T+1} - \hat{y}_{T+1}(T)) \\ &= \hat{y}_{T+1}(T) + \lambda e_{T+1}(1),\end{aligned} \tag{4.28}$$

where $e_{T+1}(1) = y_{T+1} - \hat{y}_{T+1}(T)$ is called the one-step-ahead forecast or prediction error. The interpretation of Eq. (4.28) makes it easier to understand the forecasting process using exponential smoothing: our forecast for the next observation is simply our previous forecast for the current observation plus a fraction of the forecast error we made in forecasting the current observation. The fraction in this summation is determined by λ. Hence how fast our forecast will react to the forecast error depends on the discount factor. A large discount factor will lead to fast reaction to the forecast error but it may also make our forecast react fast to random fluctuations. This once again brings up the issue of the choice of the discount factor.

Choice of λ We will define the sum of the squared one-step-ahead forecast errors as

$$SS_E(\lambda) = \sum_{t=1}^{T} e_t^2(1). \tag{4.29}$$

For a given historic data, we can in general calculate SS_E values for various values of λ and pick the value of λ that gives the smallest sum of the squared forecast errors.

Prediction Intervals Another issue in forecasting is the uncertainty associated with it. That is, we may be interested not only in the "point estimates" but also in the quantification of the prediction uncertainty. This is usually achieved by providing the prediction intervals that are expected at a specific confidence level to contain the future observations. Calculations of the prediction intervals will require the estimation of the variance of the forecast errors. We will discuss two different techniques in estimating prediction error variance in Section 4.6.3. For the constant process, the 100 $(1 - \alpha/2)$ percent prediction intervals for any lead time τ are given as

$$\tilde{y}_T \pm Z_{\alpha/2}\hat{\sigma}_e,$$

where $\tilde{y}_T$ is the first-order exponential smoother, $Z_{\alpha/2}$ is the $100(1 - \alpha/2)$ percentile of the standard normal distribution, and $\hat{\sigma}_e$ is the estimate of the standard deviation of the forecast errors.

It should be noted that the prediction interval is constant for all lead times. This of course can be (and probably is in most cases) quite unrealistic. As it will be more likely that the process goes through some changes as time goes on, we would correspondingly expect to be less and less "sure" about our predictions for large lead times (or large τ values). Hence we would anticipate prediction intervals that are getting wider and wider for increasing lead times. We propose a remedy for this in Section 4.6.3. We will discuss this issue further in Chapter 6.

Example 4.4 We are interested in the average speed on a specific stretch of a highway during nonrush hours. For the past year and a half (78 weeks), we have available weekly averages of the average speed in miles/hour between 10 AM and 3 PM. The data are given in Table 4.5. Figure 4.18 shows that the time series data follow a constant process. To smooth out the excessive variation, however, first-order exponential smoothing can be used. The "best" smoothing constant can be determined by finding the smoothing constant value that minimizes the sum of the squared one-step-ahead prediction errors.

The sum of the squared one-step-ahead prediction errors for various λ values is given in Table 4.6. Furthermore, Figure 4.19 shows that the minimum SS_E is obtained for $\lambda = 0.4$.

Let us assume that we are also asked to make forecasts for the next 12 weeks at week 78. Figure 4.20 shows the smoothed values for the first 78 weeks together with the forecasts for weeks 79–90 with prediction intervals. It also shows the actual weekly speed during that period. Note that since the constant process is assumed, the forecasts for the next 12 weeks are the same. Similarly, the prediction intervals are constant for that period.

TABLE 4.5 The Weekly Average Speed During Nonrush Hours

Week	Speed	Week	Speed	Week	Speed	Week	Speed
1	47.12	26	46.74	51	45.71	76	45.69
2	45.01	27	46.62	52	43.84	77	44.59
3	44.69	28	45.31	53	45.09	78	43.45
4	45.41	29	44.69	54	44.16	79	44.75
5	45.45	30	46.39	55	46.21	80	45.46
6	44.77	31	43.79	56	45.11	81	43.73
7	45.24	32	44.28	57	46.16	82	44.15
8	45.27	33	46.04	58	46.50	83	44.05
9	46.93	34	46.45	59	44.88	84	44.83
10	47.97	35	46.31	60	45.68	85	43.93
11	45.27	36	45.65	61	44.40	86	44.40
12	45.10	37	46.28	62	44.17	87	45.25
13	43.31	38	44.11	63	45.18	88	44.80
14	44.97	39	46.00	64	43.73	89	44.75
15	45.31	40	46.70	65	45.14	90	44.50
16	45.23	41	47.84	66	47.98	91	45.12
17	42.92	42	48.24	67	46.52	92	45.28
18	44.99	43	45.59	68	46.89	93	45.15
19	45.12	44	46.56	69	46.01	94	46.24
20	46.67	45	45.02	70	44.98	95	46.15
21	44.62	46	43.67	71	45.76	96	46.57
22	45.11	47	44.53	72	45.38	97	45.51
23	45.18	48	44.37	73	45.33	98	46.98
24	45.91	49	44.62	74	44.07	99	46.64
25	48.39	50	46.71	75	44.02	100	44.31

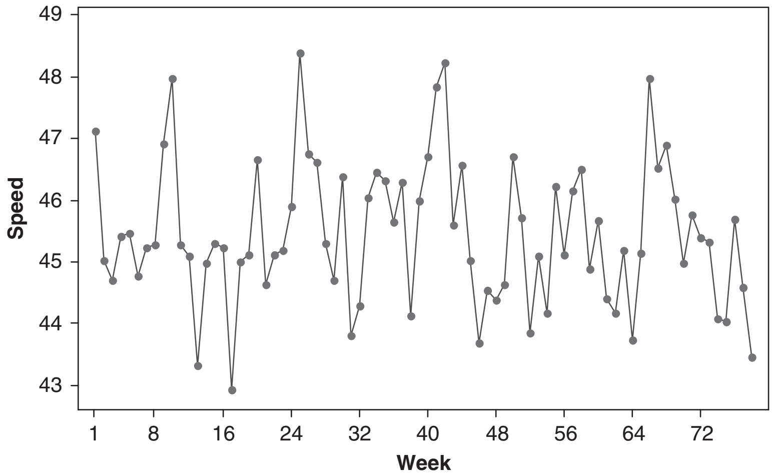


FIGURE 4.18 The weekly average speed during nonrush hours.

TABLE 4.6 SS_E for Different λ Values for the Average Speed Data

λ		0.1		0.2		0.3		0.4		0.5		0.9	
Week	Speed	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$	Forecast	$e(t)$
1	47.12	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00	47.12	0.00
2	45.01	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.11	47.12	−2.U	47.12	−2.11
3	44.69	46.91	−2.22	46.70	−2.01	46.49	−1.80	46.28	−1.59	46.07	−1.38	45.22	−0.53
4	45.41	46.69	−1.28	46.30	−0.89	45.95	−0.54	45.64	−0.23	45.38	0.03	44.74	0.67
5	45.45	46.56	−1.11	46.12	−0.67	45.79	−0.34	45.55	−0.10	45.39	0.06	45.34	0.11
6	44.77	46.45	−1.68	45.99	−1.22	45.69	−0.92	45.51	−0.74	45.42	−0.65	45.44	−0.67
7	45.24	46.28	−1.04	45.74	−0.50	45.41	−0.17	45.21	0.03	45.10	0.14	44.84	0.40
8	45.27	46.18	−0.91	45.64	−0.37	45.36	−0.09	45.22	0.05	45.17	0.10	45.20	0.07
9	46.93	46.09	0.84	45.57	1.36	45.33	1.60	45.24	1.69	45.22	1.71	45.26	1.67
10	47.97	46.17	1.80	45.84	2.13	45.81	2.16	45.92	2.05	46.07	1.90	46.76	1.21
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
75	44.02	45.42	−1.40	45.30	−1.28	45.12	−1.10	44.93	−0.91	44.75	−0.73	44.20	−0.18
76	45.69	45.28	0.41	45.05	0.64	44.79	0.90	44.56	1.13	44.39	1.30	44.04	1.65
77	44.59	45.32	−0.73	45.18	−0.59	45.06	−0.47	45.01	−0.42	45.04	−0.45	45.52	−0.93
78	43.45	45.25	−1.80	45.06	−1.61	44.92	−1.47	44.84	−1.39	44.81	−1.36	44.68	−1.23
SS_E			124.14		118.88		117.27		116.69		116.95		128.98

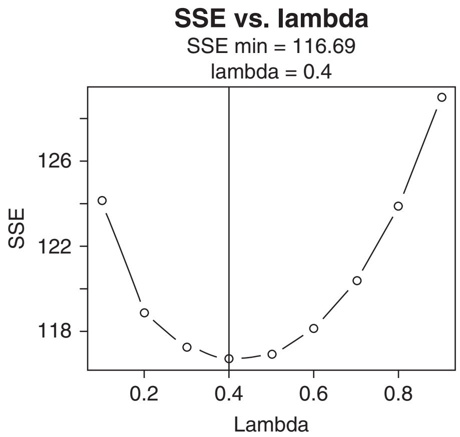


FIGURE 4.19 Plot of SS_E for various λ values for average speed data.

4.6.2 Linear Trend Process

The t-step-ahead forecast for the linear trend model is given by

$$\begin{aligned}\hat{y}_{T+\tau}(T) &= \hat{\beta}_{0,T} + \hat{\beta}_{1,T}(T+\tau) \\ &= \hat{\beta}_{0,T} + \hat{\beta}_{1,T}T + \hat{\beta}_{1,T}\tau \\ &= \hat{y}_T + \hat{\beta}_{1,T}\tau.\end{aligned} \tag{4.30}$$

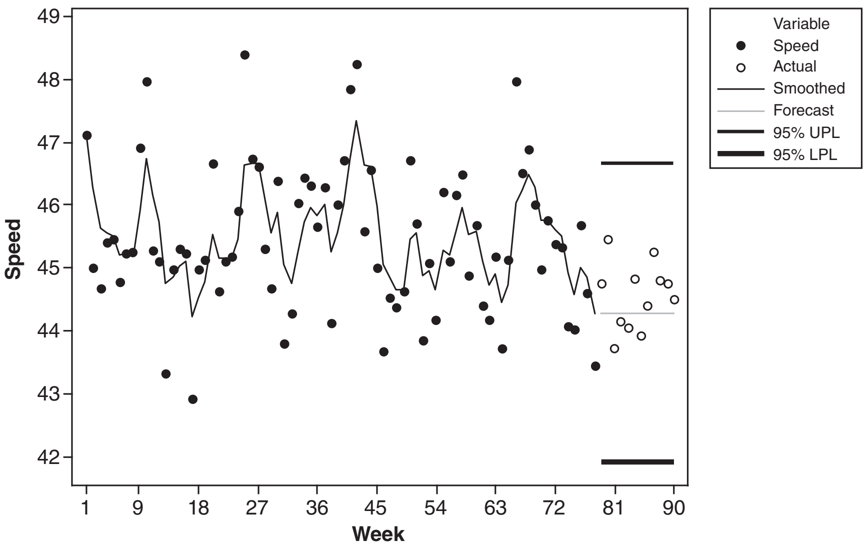


FIGURE 4.20 Forecasts for the weekly average speed data for weeks 79–90.

In terms of the exponential smoothers, we can rewrite Eq. (4.30) as

$$\begin{aligned}\hat{y}_{T+\tau}(\tau) &= \left(2\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) + \tau\frac{\lambda}{1-\lambda}\left(\tilde{y}_T^{(1)} - \tilde{y}_T^{(2)}\right) \\ &= \left(2 + \frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(1)} - \left(1 + \frac{\lambda}{1-\lambda}\tau\right)\tilde{y}_T^{(2)}.\end{aligned} \tag{4.31}$$

It should be noted that the predictions for the trend model depend on the lead time and, as opposed to the constant model, will be different for different lead times. As we collect more data, we can improve our forecasts by updating our parameter estimates using

$$\begin{aligned}\hat{\beta}_{0,T+1} &= \lambda(1+\lambda)y_{T+1} + (1-\lambda)^2(\hat{\beta}_{0,T} + \hat{\beta}_{1,T}) \\ \hat{\beta}_{1,T+1} &= \frac{\lambda}{(2-\lambda)}\left(\hat{\beta}_{0,T+1} - \hat{\beta}_{0,T}\right) + \frac{2(1-\lambda)}{(2-\lambda)}\hat{\beta}_{1,T}\end{aligned} \tag{4.32}$$

Subsequently, we can update our τ-step-ahead forecasts based on Eq. (4.32). As in the constant process, the discount factor, λ, can be estimated by minimizing the sum of the squared one-step-ahead forecast errors given in Eq. (4.29).

In this case, the $100(1 - \alpha/2)$ percent prediction interval for any lead time τ is

$$\left(2 + \frac{\lambda}{1-\lambda}\tau\right)\hat{y}_T^{(1)} - \left(1 + \frac{\lambda}{1-\lambda}\tau\right)\hat{y}_T^{(2)} \pm Z_{\alpha/2}\frac{c_\tau}{c_1}\hat{\sigma}_e,$$

where

$$c_i^2 = 1 + \frac{\lambda}{(2-\lambda)^3}[(10 - 14\lambda + 5\lambda^2) + 2i\lambda(4 - 3\lambda) + 2i^2\lambda^2].$$

Example 4.5 Consider the CPI data in Example 4.2. Assume that we are currently in December 2003 and would like to make predictions of the CPI for the following year. Although the data from January 1995 to December 2003 clearly exhibit a linear trend, we may still like to consider first-order exponential smoothing first. We will then calculate the "best" λ value that minimizes the sum of the squared one-step-ahead prediction errors. The predictions and prediction errors for various λ values are given in Table 4.7.

Figure 4.21 shows the sum of the squared one-step-ahead prediction errors (SS_E) for various values of λ.

TABLE 4.7 The Predictions and Prediction Errors for Various λ Values for CPI Data

		$\lambda = 0.1$		$\lambda = 0.2$		$\lambda = 0.3$			$\lambda = 0.9$		$\lambda = 0.99$	
Month-Year	CPI	Prediction	Error	Prediction	Error	Prediction	Error		Prediction	Error	Prediction	Error
Jan-1995	150.3	150.30	0.00	150.30	0.00	150.30	0.00	…	150.30	0.00	150.30	0.00
Feb-1995	150.9	150.30	0.60	150.30	0.60	150.30	0.60	…	150.30	0.60	150.30	0.60
Mar-1995	151.4	150.36	1.04	150.42	0.98	150.48	0.92	…	150.84	0.56	150.89	0.51
Apr-1995	151.9	150.46	1.44	150.62	1.28	150.76	1.14	…	151.34	0.56	151.39	0.51
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		⋮	⋮	⋮	⋮
Nov-2003	184.5	182.29	2.21	183.92	0.58	184.45	0.05	…	185.01	−0.51	185.00	−0.50
Dec-2003	184.3	182.51	1.79	184.03	0.27	184.46	−0.16	…	184.55	−0.25	184.51	−0.21
SS_E			1061.50		309.14		153.71			31.90		28.62

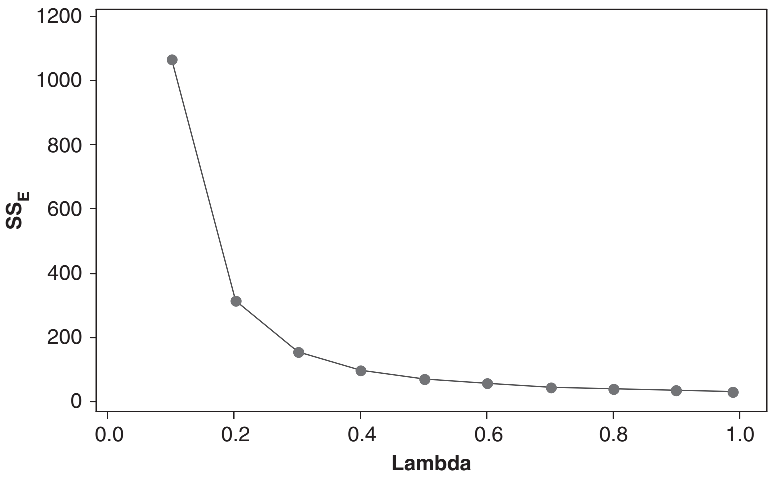


FIGURE 4.21 Scatter plot of the sum of the squared one-step-ahead prediction errors versus λ.

We notice that the SS_E keeps on getting smaller as λ gets bigger. This suggests that the data are highly autocorrelated. This can be clearly seen in the ACF plot in Figure 4.22. In fact if the "best" λ value (i.e., λ value that minimizes SS_E) turns out to be high, it may indeed be better to switch to a higher-order smoothing or use an ARIMA model as discussed in Chapter 5.

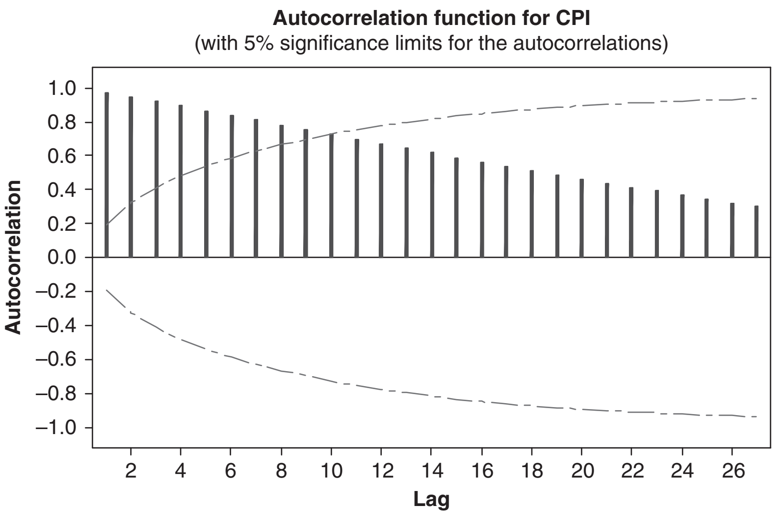


FIGURE 4.22 ACF plot for the CPI data (with 5% significance limits for the autocorrelations).

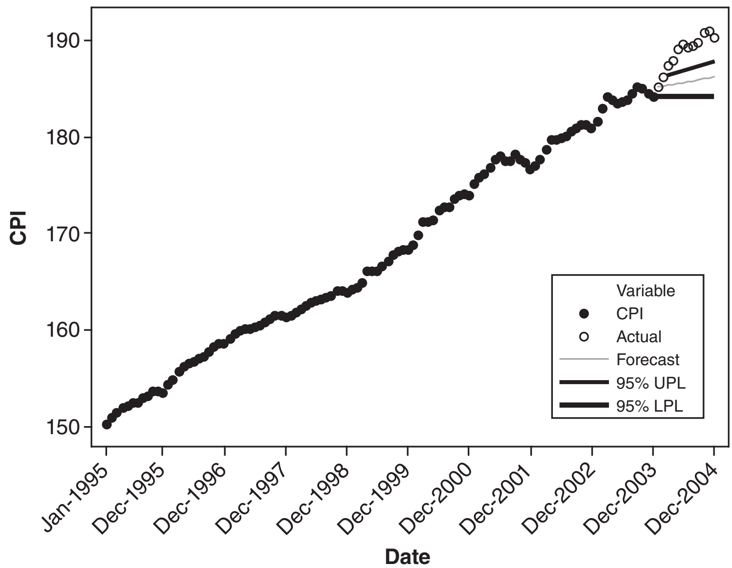


FIGURE 4.23 The 1- to 12-step-ahead forecasts of the CPI data for 2004.

Since the first-order exponential smoothing is deemed inadequate, we will now try the second-order exponential smoothing to forecast next year's monthly CPI values. Usually we have two options:

1. On December 2003, make forecasts for the entire 2004 year; that is, 1-step-ahead, 2-step-ahead, … , 12-step-ahead forecasts. For that we can use Eq. (4.30) or equivalently Eq. (4.31). Using the double exponential smoothing option in Minitab with $\lambda = 0.3$, we obtain the forecasts given in Figure 4.23.

Note that the forecasts further in the future (for the later part of 2004) are quite a bit off. To remedy this we may instead use the following strategy.

2. In December 2003, make the one-step-ahead forecast for January 2004. When the data for January 2004 becomes available, then make the one-step-ahead forecast for February 2004, and so on. We can see from Figure 4.24 that forecasts when only one-step-ahead forecasts are used and adjusted as actual data becomes available perform better than in the previous case where, for December 2003, forecasts are made for the entire following year.

The JMP software package also has an excellent forecasting capability. Table 4.8 shows output from JMP for the CPI data for double

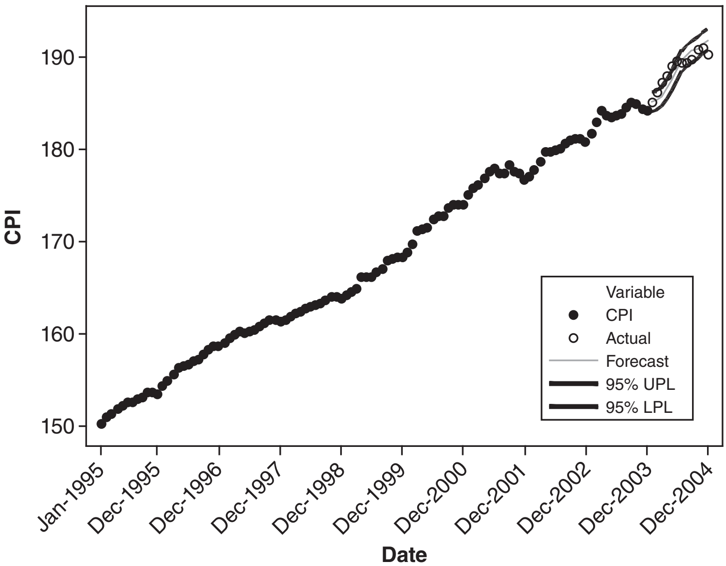
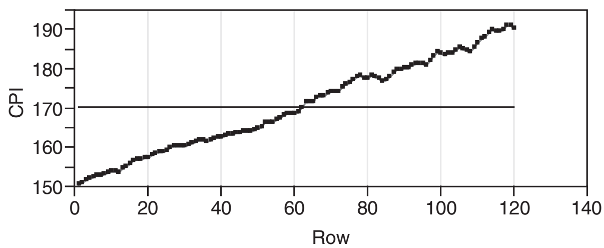


FIGURE 4.24 The one-step-ahead forecasts of the CPI data for 2004.

exponential smoothing. JMP uses the double smoothing procedure that employs a single smoothing constant. The JMP output shows the time series plot and summary statistics including the sample ACF. It also provides a sample partial ACF, which we will discuss in Chapter 5. Then an optimal smoothing constant is chosen by finding the value of λ that

TABLE 4.8 JMP Output for the CPI Data

Time series CPI



Mean	170.13167
Std	11.629323
N	120
Zero Mean ADF	8.4844029
Single Mean ADF	-0.075966
Trend ADF	-2.443095

(*continued*)

TABLE 4.8 (*Continued*)

Time series basic diagnostics

Lag	AutoCorr	Plot autocorr	Ljung-box Q	p-Value
0	1.0000		.	.
1	0.9743		116.774	<.0001
2	0.9472		228.081	<.0001
3	0.9203		334.053	<.0001
4	0.8947		435.091	<.0001
5	0.8694		531.310	<.0001
6	0.8436		622.708	<.0001
7	0.8166		709.101	<.0001
8	0.7899		790.659	<.0001
9	0.7644		867.721	<.0001
10	0.7399		940.580	<.0001
11	0.7161		1009.46	<.0001

Lag	AutoCorr	Plot autocorr	Ljung-box Q	p-Value
12	0.6924		1074.46	<.0001
13	0.6699		1135.85	<.0001
14	0.6469		1193.64	<.0001
15	0.6235		1247.84	<.0001
16	0.6001		1298.54	<.0001
17	0.5774		1345.93	<.0001
18	0.5550		1390.14	<.0001
19	0.5324		1431.24	<.0001
20	0.5098		1469.29	<.0001
21	0.4870		1504.36	<.0001
22	0.4637		1536.48	<.0001
23	0.4416		1565.91	<.0001
24	0.4205		1592.87	<.0001
25	0.4000		1617.54	0.0000

Lag	Partial	plot partial
0	1.0000	
1	0.9743	
2	−0.0396	
3	−0.0095	
4	0.0128	
5	−0.0117	
6	−0.0212	
7	−0.0379	
8	−0.0070	
9	0.0074	
10	0.0033	
11	−0.0001	
12	−0.0116	
13	0.0090	
14	−0.0224	
15	−0.0220	
16	−0.0139	
17	−0.0022	
18	−0.0089	
19	−0.0174	
20	−0.0137	
21	−0.0186	
22	−0.0234	
23	0.0074	
24	0.0030	
25	−0.0036	

TABLE 4.8 (*Continued*)

Model Comparison

Model	DF	Variance	AIC
Double (Brown) Exponential Smoothing	117	0.247119	171.05558

SBC	RSquare	-2LogLH	AIC	Rank	SBC Rank	MAPE	MAE
173.82626	0.998	169.05558	0	0	0.216853	0.376884	

Model: Double (Brown) Exponential Smoothing

Model Summary

DF	117
Sum of Squared Errors	28.9129264
Variance Estimate	0.24711903
Standard Deviation	0.49711068
Akaike's 'A' Information Criterion	171.055579
Schwarz's Bayesian Criterion	173.826263
RSquare	0.99812888
RSquare Adj	0.99812888
MAPE	0.21685285
MAE	0.37688362
-2LogLikelihood	169.055579

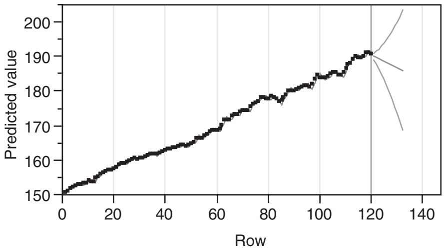
Stable Yes

Invertible Yes

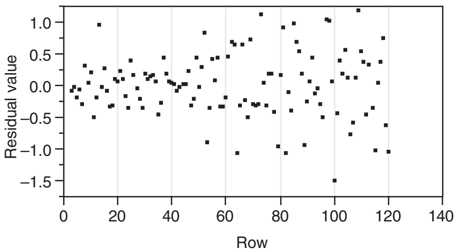
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Level Smoothing Weight	0.81402446	0.0919040	8.86	<.0001

Forecast



Residuals



(*continued*)

TABLE 4.8 (*Continued*)

Lag	AutoCorr plot autocorr	Ljung-box Q	p-Value
0	1.0000	.	.
1	0.0791	0.7574	0.3841
2	–0.3880	19.1302	<.0001
3	–0.2913	29.5770	<.0001
4	–0.0338	29.7189	<.0001
5	0.1064	31.1383	<.0001
6	0.1125	32.7373	<.0001
7	0.1867	37.1819	<.0001
8	–0.1157	38.9063	<.0001
9	–0.3263	52.7344	<.0001
10	–0.1033	54.1324	<.0001
11	0.2149	60.2441	<.0001
12	0.2647	69.6022	<.0001
13	–0.0773	70.4086	<.0001
14	0.0345	70.5705	<.0001
15	–0.1243	72.6937	<.0001
16	–0.1429	75.5304	<.0001
17	0.0602	76.0384	<.0001
18	0.1068	77.6533	<.0001
19	0.0370	77.8497	<.0001
20	–0.0917	79.0656	<.0001
21	–0.0363	79.2579	<.0001
22	–0.0995	80.7177	<.0001
23	–0.0306	80.8570	<.0001
24	0.2602	91.0544	<.0001
25	0.1728	95.6007	<.0001

Lag	Partial plot partial
0	1.0000
1	0.0791
2	–0.3967
3	–0.2592
4	–0.1970
5	–0.1435
6	–0.0775
7	0.1575
8	–0.1144
9	–0.2228
10	–0.1482

Lag	AutoCorr plot autocorr	Ljung-box Q	p-Value
11	–0.0459		
12	0.0368		
13	–0.1335		
14	0.2308		
15	–0.0786		
16	0.0050		
17	0.0390		
18	–0.0903		
19	–0.0918		
20	0.0012		
21	–0.0077		
22	–0.1935		
23	–0.0665		
24	0.1783		
25	0.0785		

minimizes the error sum of squares. The value selected is $\lambda = 0.814$. This relatively large value is not unexpected, because there is a very strong linear trend in the data and considerable autocorrelation. Values of the forecast for the next 12 periods at origin December 2004 and the associated prediction interval are also shown. Finally, the residuals from the model fit are shown along with the sample ACF and sample partial ACF plots of the residuals. The sample ACF indicates that there may be a small amount of structure in the residuals, but it is not enough to cause concern.

4.6.3 Estimation of σ_e^2

In the estimation of the variance of the forecast errors, σ_e^2, it is often assumed that the model (e.g., constant, linear trend) is correct and constant in time. With these assumptions, we have two different ways of estimating σ_e^2:

1. We already defined the one-step-ahead forecast error as $e_T(1) = y_T - \hat{y}_T(T-1)$. The idea is to apply the model to the historic data and obtain the forecast errors to calculate:

$$\begin{aligned} \hat{\sigma}_e^2 &= \frac{1}{T}\sum_{t=1}^{T} e_t^2(1) \\ &= \frac{1}{T}\sum_{t=1}^{T} (y_t - \hat{y}_t(t-1))^2 \end{aligned} \tag{4.33}$$

It should be noted that in the variance calculations the mean adjustment was not needed, since for the correct model the forecasts are unbiased; that is, the expected value of the forecast errors is 0.

As more data are collected, the variance of the forecast errors can be updated as

$$\hat{\sigma}_{eT+1}^2 = \frac{1}{T+1}\left(T\hat{\sigma}_{e,T}^2 + e_{T+1}^2(1)\right). \tag{4.34}$$

As discussed in Section 4.6.1, it may be counterintuitive to have a constant forecast error variance for all lead times. We can instead define $\sigma_e^2(\tau)$ as the τ-step-ahead forecast error variance and estimate it by

$$\hat{\sigma}_e^2(\tau) = \frac{1}{T-\tau+1}\sum_{t=\tau}^{T} e_1^2(\tau). \tag{4.35}$$

Hence the estimate in Eq. (4.35) can instead be used in the calculations of the prediction interval for the τ-step-ahead forecast.

2. For the second method of estimating σ_e^2 we will first define the *mean absolute deviation* Δ as

$$\Delta = E(|e - E(e)|) \tag{4.36}$$

and, assuming that the model is correct, calculate its estimate by

$$\hat{\Delta}_T = \delta|e_T(1)| + (1 - \delta)\hat{\Delta}_{T-1}. \tag{4.37}$$

Then the estimate of the σ_e^2 is given by

$$\hat{\sigma}_{e,T} = 1.25\hat{\Delta}_T. \tag{4.38}$$

For further details, see Montgomery et al. (1990).

4.6.4 Adaptive Updating of the Discount Factor

In the previous sections we discussed estimation of the "best" discount factor, $\hat{\lambda}$, by minimizing the sum of the squared one-step-ahead forecasts errors. However, as we have seen with the Dow Jones Index data, changes in the underlying time series model will make it difficult for the exponential smoother with fixed discount factor to follow these changes. Hence a need for monitoring and, if necessary, modifying the discount factor arises. By doing so, the discount factor will adapt to the changes in the time series model. For that we will employ the procedure originally described by Trigg and Leach (1967) for single discount factor. As an example we will consider the first-order exponential smoother and modify it as

$$\hat{y}_T = \lambda_T y_T + (1 - \lambda_T)\tilde{y}_{T-1}. \tag{4.39}$$

Please note that in Eq. (4.39), the discount factor λ_T is given as a function of time and hence it is allowed to adapt to changes in the time series model. We also define the *smoothed error* as

$$Q_T = \delta e_T(1) + (1 - \delta)Q_{T-1}, \tag{4.40}$$

where δ is a smoothing parameter.

Finally, we define the tracking signal as

$$\frac{Q_T}{\hat{\Delta}_T}, \tag{4.41}$$

where $\hat{\Delta}_T$ is given in Eq. (4.37). This ratio is expected to be close to 0 when the forecasting system performs well and to approach ± 1 as it starts to fail. In fact, Trigg and Leach (1967) suggest setting the discount factor to

$$\lambda_T = \left| \frac{Q_T}{\hat{\Delta}_T} \right| \tag{4.42}$$

Equation (4.42) will allow for automatic updating of the discount factor.

Example 4.6 Consider the Dow Jones Index from June 1999 to June 2006 given in Table 4.1. Figure 4.2 shows that the data do not exhibit a single regime of constant or linear trend behavior. Hence a single exponential smoother with adaptive discount factor as given in Eq. (4.42) can be used. Figure 4.25 shows two simple exponential smoothers for the Dow Jones Index: one with fixed $\lambda = 0.3$ and another one with adaptive updating based on the Trigg–Leach method given in Eq. (4.42).

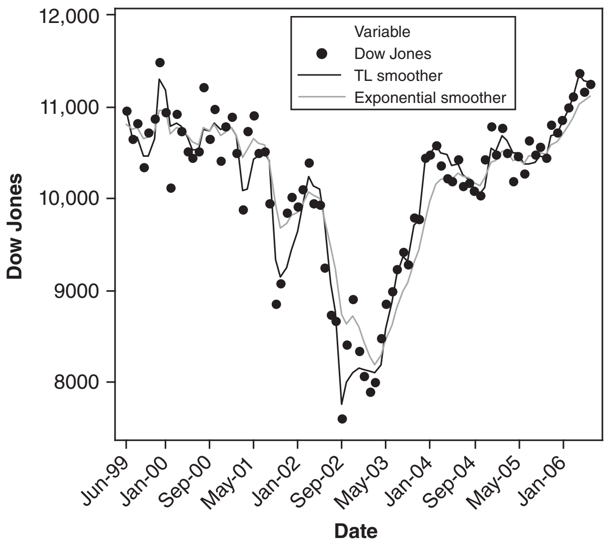


FIGURE 4.25 Time series plot of the Dow Jones Index from June 1999 to June 2006, the simple exponential smoother with $\lambda = 0.3$, and the Trigg–Leach (TL) smoother with $\delta = 0.3$.

TABLE 4.9 The Trigg–Leach Smoother for the Dow Jones Index

Date	Dow Jones	Smoothed	λ	Error	Q_t	D_t
Jun-99	10,970.8	10,970.8	1		0	0
Jul-99	10,655.2	10,655.2	1	−315.6	−94.68	94.68
Aug-99	10,829.3	10,675.835	0.11853	174.1	−14.046	118.506
Sep-99	10,337	10,471.213	0.6039	−338.835	−111.483	184.605
Oct-99	10,729.9	10,471.753	0.00209	258.687	−0.43178	206.83
⋮	⋮	⋮	⋮	⋮	⋮	⋮
May-06	11,168.3	11,283.962	0.36695	−182.705	68.0123	185.346
Jun-06	11,247.9	11,274.523	0.26174	−36.0619	36.79	140.561

This plot shows that a better smoother can be obtained by making automatic updates to the discount factor. The calculations for the Trigg–Leach smoother are given in Table 4.9.

The adaptive smoothing procedure suggested by Trigg and Leach is a useful technique. For other approaches to adaptive adjustment of exponential smoothing parameters, see Chow (1965), Roberts and Reed (1969), and Montgomery (1970).

4.6.5 Model Assessment

If the forecast model performs as expected, the forecast errors should not exhibit any pattern or structure; that is, they should be uncorrelated. Therefore it is always a good idea to verify this. As noted in Chapter 2, we can do so by calculating the sample ACF of the forecast errors from

$$rk = \frac{\sum_{t=k}^{T-1} [e_t(1) - \bar{e}] \left[e_{t-k}(1) - \bar{e}\right]}{\sum_{T=0}^{T-1} [e_t(1) - \bar{e}]^2}, \tag{4.43}$$

where

$$\bar{e} = \frac{1}{n} \sum_{t=1}^{T} e_t(1).$$

If the one-step-ahead forecast errors are indeed uncorrelated, the sample autocorrelations for any lag k should be around 0 with a standard error $1/\sqrt{T}$. Hence a sample autocorrelation for any lag k that lies outside the $\pm 2/\sqrt{T}$ limits will require further investigation of the model.

4.7 EXPONENTIAL SMOOTHING FOR SEASONAL DATA

Some time series data exhibit cyclical or seasonal patterns that cannot be effectively modeled using the polynomial model in Eq. (4.25). Several approaches are available for the analysis of such data. In this chapter we will discuss exponential smoothing techniques that can be used in modeling seasonal time series. The methodology we will focus on was originally introduced by Holt (1957) and Winters (1960) and is generally known as Winters' method, where a seasonal adjustment is made to the linear trend model. Two types of adjustments are suggested—additive and multiplicative.

4.7.1 Additive Seasonal Model

Consider the US clothing sales data given in Figure 4.26. Clearly, for certain months of every year we have high (or low) sales. Hence we can conclude that the data exhibit seasonality. The data also exhibit a linear trend as the sales tend to get higher for the same month as time goes on. As the final observation, we note that the amplitude of the seasonal pattern, that is, the range of the periodic behavior within a year, remains more or

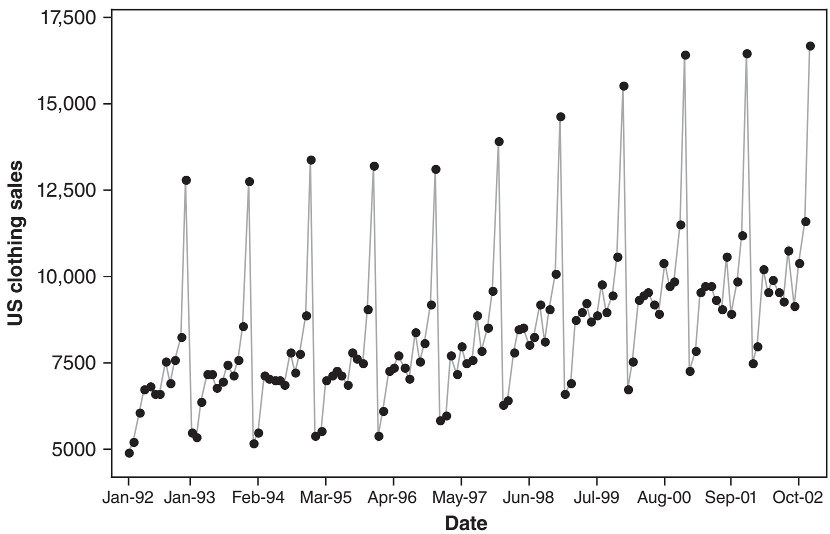


FIGURE 4.26 Time series plot of US clothing sales from January 1992 to December 2003.

less constant in time and remains independent of the average level within a year.

We will for this case assume that the seasonal time series can be represented by the following model:

$$y_t = L_t + S_t + \varepsilon_t, \tag{4.44}$$

where L_t represents the level or linear trend component and can in turn be represented by $\beta_0 + \beta_1 t$; S_t represents the seasonal adjustment with $S_t = S_{t+s} = S_{t+2s} = \dots$ for $t = 1, \dots, s-1$, where s is the length of the season (period) of the cycles; and the ε_t are assumed to be uncorrelated with mean 0 and constant variance σ_ε^2. Sometimes the level is called the *permanent component*. One usual restriction on this model is that the seasonal adjustments add to zero during one season,

$$\sum_{t=1}^{s} S_t = 0. \tag{4.45}$$

In the model given in Eq. (4.44), for forecasting the future observations, we will employ first-order exponential smoothers with different discount factors. The procedure for updating the parameter estimates once the current observation y_T is obtained is as follows.

Step 1. Update the estimate of L_T using

$$\hat{L}_T = \lambda_1 (y_T - \hat{S}_{T-s}) + (1 - \lambda_1)\left(\hat{L}_{T-1} + \hat{\beta}_{1,T-1}\right), \tag{4.46}$$

where $0 < \lambda_1 < 1$. It should be noted that in Eq. (4.46), the first part can be seen as the "current" value for L_T and the second part as the forecast of L_T based on the estimates at $T-1$.

Step 2. Update the estimate of β_1 using

$$\hat{\beta}_{1,T} = \lambda_2 (\hat{L}_T - \hat{L}_{T-1}) + (1 - \lambda_2)\hat{\beta}_{1,T-1}, \tag{4.47}$$

where $0 < \lambda_2 < 1$. As in Step 1, the estimate of β_1 in Eq. (4.47) can be seen as the linear combination of the "current" value of β_1 and its "forecast" at $T-1$.

Step 3. Update the estimate of S_t using

$$\hat{S}_T = \lambda_3 (y_T - \hat{L}_T) + (1 - \lambda_3)\hat{S}_{T-s}, \tag{4.48}$$

where $0 < \lambda_3 < 1$.

Step 4. Finally, the τ-step-ahead forecast, $\hat{y}_{T+\tau}(T)$, is

$$\hat{y}_{T+\tau}(T) = \hat{L}_T + \hat{\beta}_{1,T}\tau + \hat{S}_T(\tau - s). \tag{4.49}$$

As before, estimating the initial values of the exponential smoothers is important. For a given set of historic data with n seasons (hence ns observations), we can use the least squares estimates of the following model:

$$y_t = \beta_0 + \beta_1 t + \sum_{i=1}^{s-1} \gamma_i (I_{t,i} - I_{t,s}) + \varepsilon_t, \tag{4.50}$$

where

$$I_{t,i} = \begin{cases} 1, & t = i, i+s, i+2s, \ldots \\ 0, & \text{otherwise} \end{cases}. \tag{4.51}$$

The least squares estimates of the parameters in Eq. (4.50) are used to obtain the initial values as

$$\begin{aligned}
&\hat{\beta}_{0,0} = \hat{L}_0 = \hat{\beta}_0 \\
&\hat{\beta}_{1,0} = \hat{\beta}_1 \\
&\hat{S}_{j-s} = \hat{Y}_{\mathrm{j}} \quad \text{for } 1 \le j \le \mathrm{s} - 1 \\
&\hat{S}_0 = -\sum_{j=1}^{s-1} \hat{y}_j
\end{aligned}$$

These are initial values of the model parameters at the *original* origin of time, $t = 0$. To make forecasts from the correct origin of time the permanent component must be shifted to time T by computing $\hat{L}_T = \hat{L}_0 + ns\hat{\beta}_1$. Alternatively, one could smooth the parameters using equations (4.46)–(4.48) for time periods $t = 1, 2, \ldots, \mathrm{T}$.

Prediction Intervals As in the nonseasonal smoothing case, the calculations of the prediction intervals would require an estimate for the prediction error variance. The most common approach is to use the relationship between the exponential smoothing techniques and the ARIMA models of Chapter 5 as discussed in Section 4.8, and estimate the prediction error variance accordingly. It can be shown that the seasonal exponential smoothing using the three parameter Holt–Winters method is optimal for an ARIMA $(0, 1, s+1) \times (0, 1, 0)_s$, process, where s represents the length of the period of the seasonal cycles. For further details, see Yar and Chatfield (1990) and McKenzie (1986).

An alternate approach is to recognize that the additive seasonal model is just a linear regression model and to use the ordinary least squares (OLS) regression procedure for constructing prediction intervals as discussed in Chapter 3. If the errors are correlated, the regression methods for autocorrelated errors could be used instead of OLS.

Example 4.7 Consider the clothing sales data given in Table 4.10. To obtain the smoothed version of this data, we can use the Winters' method option in Minitab. Since the amplitude of the seasonal pattern is constant over time, we decide to use the additive model. Two issues we have encountered in previous exponential smoothers have to be addressed in this case as well—initial values and the choice of smoothing constants. Similar recommendations as in the previous exponential smoothing options can also be made in this case. Of course, the choice of the smoothing constant, in particular, is a bit more concerning since it involves the estimation of three smoothing constants. In this example, we follow our usual recommendation and choose smoothing constants that are all equal to 0.2. For more complicated cases, we recommend seasonal ARIMA models, which we will discuss in Chapter 5.

Figure 4.27 shows the smoothed version of the seasonal clothing sales data. To use this model for forecasting, let us assume that we are currently in December 2002 and we are asked to make forecasts for the following year. Figure 4.28 shows the forecasted sales for 2003 together with the actual data and the 95% prediction limits. Note that the forecast for December 2003 is the 12-step-ahead forecast made in December 2002. Even though the forecast is made further in the future, it still performs well since in the "seasonal" sense it is in fact a one-step-ahead forecast.

4.7.2 Multiplicative Seasonal Model

If the amplitude of the seasonal pattern is proportional to the average level of the seasonal time series, as in the liquor store sales data given in Figure 4.29, the following multiplicative seasonal model will be more appropriate:

$$y_t = L_t S_t + \varepsilon_t, \tag{4.52}$$

where L_t once again represents the permanent component (i.e., $\beta_0 + \beta_1 t$); S_t represents the seasonal adjustment with $S_t = S_{t+s} = S_{t+2s} = \cdots$ for

TABLE 4.10 US Clothing Sales from January 1992 to December 2003

Date	Sales	Date	Sales	Date	Sales	Date	Sales	Date	Sales
Jan-92	4889	Aug-94	7824	Mar-97	7695	Oct-99	9481	May-02	9906
Feb-92	5197	Sep-94	7229	Apr-97	7161	Nov-99	10577	Jun-02	9530
Mar-92	6061	Oct-94	7772	May-97	7978	Dec-99	15552	Jul-02	9298
Apr-92	6720	Nov-94	8873	Jun-97	7506	Jan-00	6726	Aug-02	10,755
May-92	6811	Dec-94	13397	Jul-97	7602	Feb-00	7514	Sep-02	9128
Jun-92	6579	Jan-95	5377	Aug-97	8877	Mar-00	9330	Oct-02	10,408
Jul-92	6598	Feb-95	5516	Sep-97	7859	Apr-00	9472	Nov-02	11,618
Aug-92	7536	Mar-95	6995	Oct-97	8500	May-00	9551	Dec-02	16,721
Sep-92	6923	Apr-95	7131	Nov-97	9594	Jun-00	9203	Jan-03	7891
Oct-92	7566	May-95	7246	Dec-97	13952	Jul-00	8910	Feb-03	7892
Nov-92	8257	Jun-95	7140	Jan-98	6282	Aug-00	10378	Mar-03	9874
Dec-92	12,804	Jul-95	6863	Feb-98	6419	Sep-00	9731	Apr-03	9920
Jan-93	5480	Aug-95	7790	Mar-98	7795	Oct-00	9868	May-03	10,431
Feb-93	5322	Sep-95	7618	Apr-98	8478	Nov-00	11512	Jun-03	9758
Mar-93	6390	Oct-95	7484	May-98	8501	Dec-00	16422	Jul-03	10,003
Apr-93	7155	Nov-95	9055	Jun-98	8044	Jan-01	7263	Aug-03	11,055

(*continued*)

TABLE 4.10 (*Continued*)

Date	Sales	Date	Sales	Date	Sales	Date	Sales	Date	Sales
May-93	7175	Dec-95	13,201	Jul-98	8272	Feb-01	7866	Sep-03	9941
Jun-93	6770	Jan-96	5375	Aug-98	9189	Mar-01	9535	Oct-03	10,763
Jul-93	6954	Feb-96	6105	Sep-98	8099	Apr-01	9710	Nov-03	12058
Aug-93	7438	Mar-96	7246	Oct-98	9054	May-01	9711	Dec-03	17535
Sep-93	7144	Apr-96	7335	Nov-98	10,093	Jun-01	9324		
Oct-93	7585	May-96	7712	Dec-98	14668	Jul-01	9063		
Nov-93	8558	Jun-96	7337	Jan-99	6617	Aug-01	10,584		
Dec-93	12,753	Jul-96	7059	Feb-99	6928	Sep-01	8928		
Jan-94	5166	Aug-96	8374	Mar-99	8734	Oct-01	9843		
Feb-94	5464	Sep-96	7554	Apr-99	8973	Nov-01	11,211		
Mar-94	7145	Oct-96	8087	May-99	9237	Dec-01	16,470		
Apr-94	7062	Nov-96	9180	Jun-99	8689	Jan-02	7508		
May-94	6993	Dec-96	13109	Jul-99	8869	Feb-02	8002		
Jun-94	6995	Jan-97	5833	Aug-99	9764	Mar-02	10,203		
Jul-94	6886	Feb-97	5949	Sep-99	8970	Apr-02	9548		

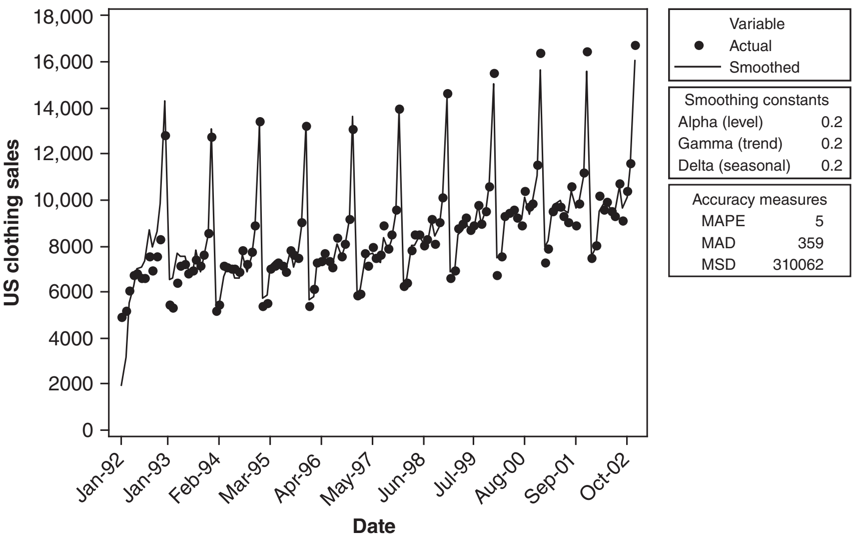


FIGURE 4.27 Smoothed data for the US clothing sales from January 1992 to December 2003 using the additive model.

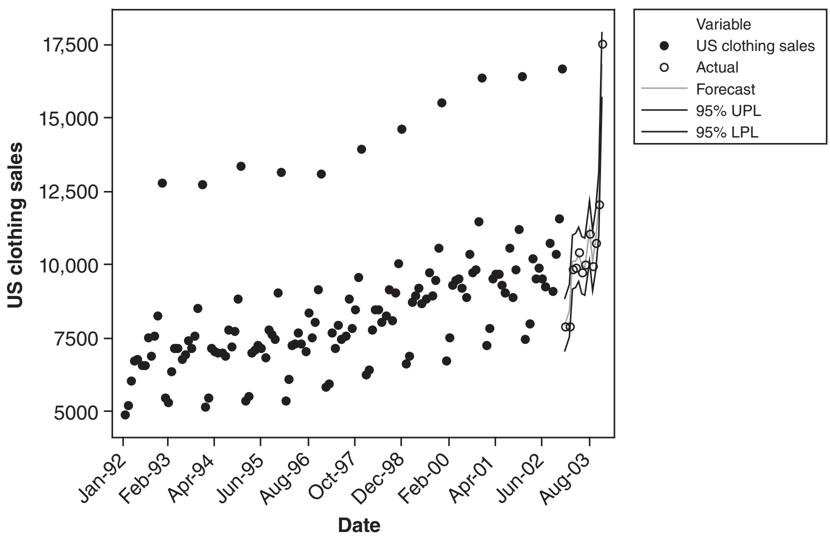


FIGURE 4.28 Forecasts for 2003 for the US clothing sales.

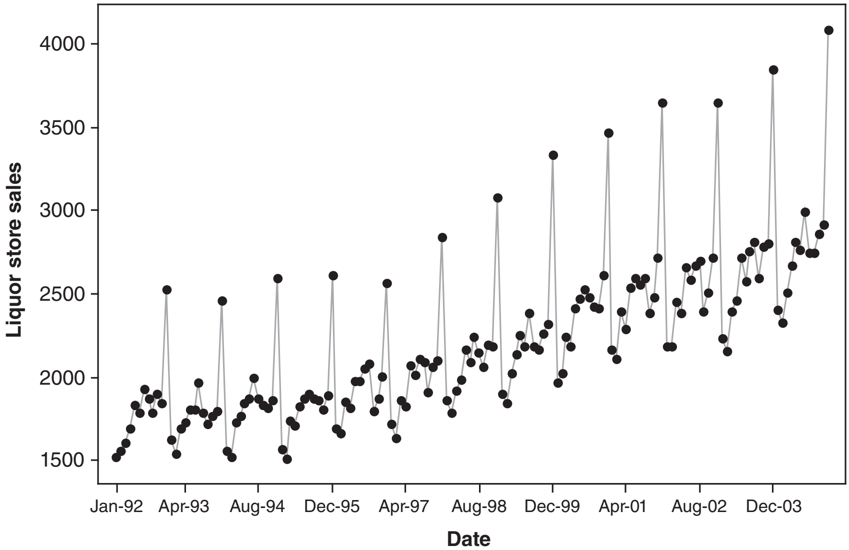


FIGURE 4.29 Time series plot of liquor store sales data from January 1992 to December 2004.

$t = i, \ldots, s - 1$, where s is the length of the period of the cycles; and the ε_t are assumed to be uncorrelated with mean 0 and constant variance σ_ε^2. The restriction for the seasonal adjustments in this case becomes

$$\sum_t^s S_t = s. \tag{4.53}$$

As in the additive model, we will employ three exponential smoothers to estimate the parameters in Eq. (4.52).

Step 1. Update the estimate of L_T using

$$\hat{L}_T = \lambda_1 \frac{y_T}{\hat{S}_{T-s}} + (1 - \lambda_1)(\hat{L}_{T-1} + \hat{\beta}_{1,T-1}), \tag{4.54}$$

where $0 < \lambda_1 < 1$. Similar interpretation as in the additive model can be made for the exponential smoother in Eq. (4.54).

Step 2. Update the estimate of β_1 using

$$\hat{\beta}_{1,T} = \lambda_2(\hat{L}_T - \hat{L}_{T-1}) + (1 - \lambda_2)\hat{\beta}_{1,T-1}, \tag{4.55}$$

where $0 < \lambda_2 < 1$.

Step 3. Update the estimate of S_t using

$$\hat{S}_T = \lambda_3 \frac{y_T}{\hat{L}_T} + (1 - \lambda_3)\hat{S}_{T-s}, \tag{4.56}$$

where $0 < \lambda_3 < 1$.

Step 4. The τ-step-ahead forecast, $\hat{y}_{T+\tau}(T)$, is

$$\hat{y}_{T+\tau}(T) = (\hat{L}_T + \hat{\beta}_{1,T}\tau)\hat{S}_T(\tau - s). \tag{4.57}$$

It will almost be necessary to obtain starting values of the model parameters. Suppose that a record consisting of n seasons of data is available. From this set of historical data, the initial values, $\hat{\beta}_{0,0}$, $\hat{\beta}_{1,0}$, and $\hat{S}_0$, can be calculated as

$$\hat{\beta}_{0,0} = \hat{L}_0 = \frac{\bar{y}_n - \bar{y}_1}{(n-1)s},$$

where

$$\bar{y}_i = \frac{1}{s} \sum_{t=(i-1)s+1}^{is} y_t$$

and

$$\hat{\beta}_{1,0} = \bar{y}_1 - \frac{s}{2}\hat{\beta}_{0,0}$$

$$\hat{S}_{j-s} = s \frac{\hat{S}_j^*}{\sum_{i=1}^{s} \hat{S}_i^*} \text{for } 1 \leq j \leq s,$$

where

$$\hat{S}_j^* = \frac{1}{n} \sum_{t=1}^{\mathrm{n}} \frac{y_{(t-1)s+j}}{\bar{y}_t - ((s+1)/2 - j)\,\hat{\beta}_0}.$$

For further details, please see Montgomery et al. (1990) and Abraham and Ledolter (1983).

Prediction Intervals Constructing prediction intervals for the multiplicative model is much harder than the additive model as the former is nonlinear. Several authors have considered this problem, including Chatfield and Yar (1991), Sweet (1985), and Gardner (1988). Chatfield and Yar (1991) propose an empirical method in which the length of the prediction interval depends on the point of origin of the forecast and may decrease in length near the low points of the seasonal cycle. They also discuss the case where the error is assumed to be proportional to the seasonal effect rather than constant, which is the standard assumption in Winters' method. Another approach would be to obtain a "linearized" version of Winters' model by expanding it in a first-order Taylor series and use this to find an approximate variance of the predicted value (statisticians call this the delta method). Then this prediction variance could be used to construct prediction intervals much as is done in the linear regression model case.

Example 4.8 Consider the liquor store data given in Table 4.11. In Figure 4.29, we can see that the amplitude of the periodic behavior gets larger as the average level of the seasonal data gets larger due to a linear trend. Hence the multiplicative model will be more appropriate. Figures 4.30 and 4.31 show the smoothed data with additive and multiplicative models, respectively. Based on the performance of the smoothers, it should therefore be clear that the multiplicative model should indeed be preferred.

As for forecasting using the multiplicative model, we can assume as usual that we are currently in December 2003 and are asked to forecast the sales in 2004. Figure 4.32 shows the forecasts together with the actual values and the prediction intervals.

4.8 EXPONENTIAL SMOOTHING OF BIOSURVEILLANCE DATA

Bioterrorism is the use of biological agents in a campaign of aggression. The use of biological agents in warfare is not new; many centuries ago plague and other contagious diseases were employed as weapons. Their use today is potentially catastrophic, so medical and public health officials are designing and implementing biosurveillance systems to monitor populations for potential disease outbreaks. For example, public health officials collect syndrome data from sources such as hospital emergency rooms, outpatient clinics, and over-the-counter medication sales to detect disease outbreaks, such as the onset of the flu season. For an excellent and highly readable introduction to statistical techniques for biosurveillance and syndromic surveillance, see Fricker (2013). Monitoring of syndromic data is also a type of epidemiologic surveillance in a biosurveillance process,

TABLE 4.11 Liquor Store Sales from January 1992 to December 2004

Date	Sales	Date	Sales	Date	Sales	Date	Sales	Date	Sales
Jan-92	1519	Aug-94	1870	Mar-97	1862	Oct-99	2264	May-02	2661
Feb-92	1551	Sep-94	1834	Apr-97	1826	Nov-99	2321	Jun-02	2579
Mar-92	1606	Oct-94	1817	May-97	2071	Dec-99	3336	Jul-02	2667
Apr-92	1686	Nov-94	1857	Jun-97	2012	Jan-00	1963	Aug-02	2698
May-92	1834	Dec-94	2593	Jul-97	2109	Feb-00	2022	Sep-02	2392
Jun-92	1786	Jan-95	1565	Aug-97	2092	Mar-00	2242	Oct-02	2504
Jul-92	1924	Feb-95	1510	Sep-97	1904	Apr-00	2184	Nov-02	2719
Aug-92	1874	Mar-95	1736	Oct-97	2063	May-00	2415	Dec-02	3647
Sep-92	1781	Apr-95	1709	Nov-97	2096	Jun-00	2473	Jan-03	2228
Oct-92	1894	May-95	1818	Dec-97	2842	Jul-00	2524	Feb-03	2153
Nov-92	1843	Jun-95	1873	Jan-98	1863	Aug-00	2483	Mar-03	2395
Dec-92	2527	Jul-95	1898	Feb-98	1786	Sep-00	2419	Apr-03	2460
Jan-93	1623	Aug-95	1872	Mar-98	1913	Oct-00	2413	May-03	2718
Feb-93	1539	Sep-95	1856	Apr-98	1985	Nov-00	2615	Jun-03	2570
Mar-93	1688	Oct-95	1800	May-98	2164	Dec-00	3464	Jul-03	2758
Apr-93	1725	Nov-95	1892	Jun-98	2084	Jan-01	2165	Aug-03	2809

(continued)

TABLE 4.11 (***Continued***)

Date	Sales	Date	Sales	Date	Sales	Date	Sales	Date	Sales
May-93	1807	Dec-95	2616	Jul-98	2237	Feb-01	2107	Sep-03	2597
Jun-93	1804	Jan-96	1690	Aug-98	2146	Mar-01	2390	Oct-03	2785
Jul-93	1962	Feb-96	1662	Sep-98	2058	Apr-01	2292	Nov-03	2803
Aug-93	1788	Mar-96	1849	Oct-98	2193	May-01	2538	Dec-03	3849
Sep-93	1717	Apr-96	1810	Nov-98	2186	Jun-01	2596	Jan-04	2406
Oct-93	1769	May-96	1970	Dec-98	3082	Jul-01	2553	Feb-04	2324
Nov-93	1794	Jun-96	1971	Jan-99	1897	Aug-01	2590	Mar-04	2509
Dec-93	2459	Jul-96	2047	Feb-99	1838	Sep-01	2384	Apr-04	2670
Jan-94	1557	Aug-96	2075	Mar-99	2021	Oct-01	2481	May-04	2809
Feb-94	1514	Sep-96	1791	Apr-99	2136	Nov-01	2717	Jun-04	2764
Mar-94	1724	Oct-96	1870	May-99	2250	Dec-01	3648	Jul-04	2995
Apr-94	1769	Nov-96	2003	Jun-99	2186	Jan-02	2182	Aug-04	2745
May-94	1842	Dec-96	2562	Jul-99	2383	Feb-02	2180	Sep-04	2742
Jun-94	1869	Jan-97	1716	Aug-99	2182	Mar-02	2447	Oct-04	2863
Jul-94	1994	Feb-97	1629	Sep-99	2169	Apr-02	2380	Nov-04	2912
								Dec-04	4085

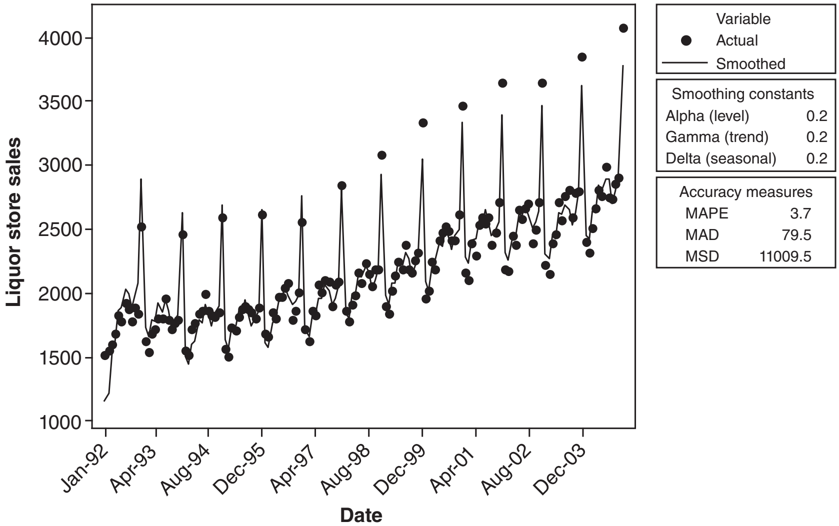


FIGURE 4.30 Smoothed data for the liquor store sales from January 1992 to December 2004 using the additive model.

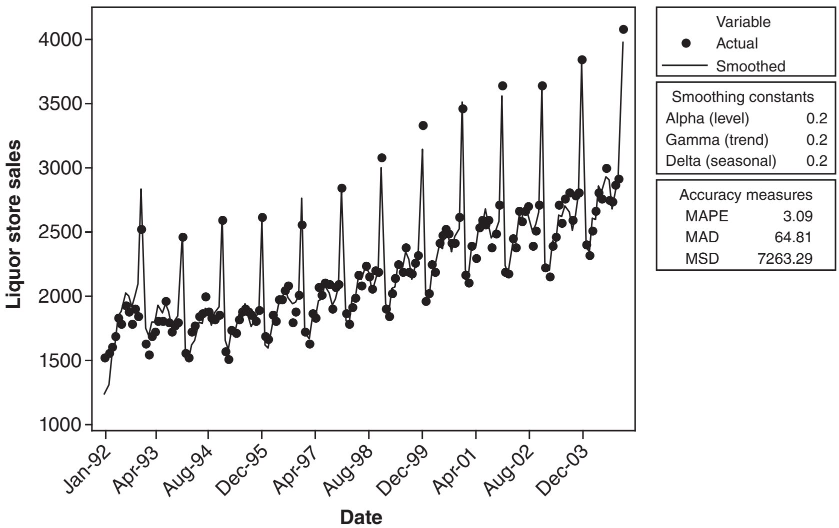


FIGURE 4.31 Smoothed data for the liquor store sales from January 1992 to December 2004 using the multiplicative model.

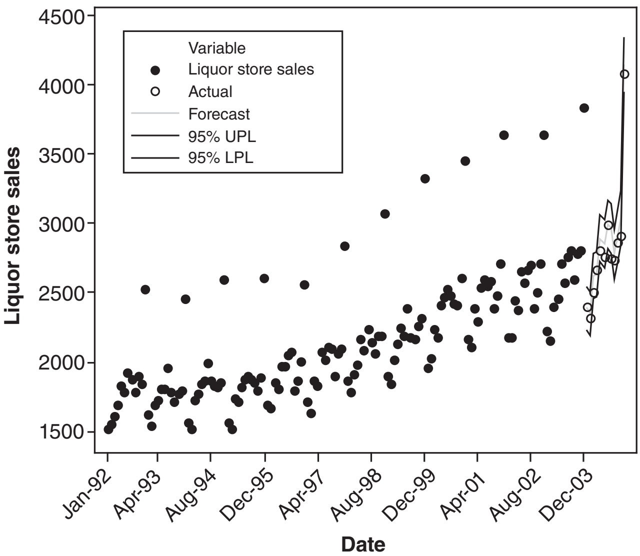


FIGURE 4.32 Forecasts for the liquor store sales for 2004 using the multiplicative model.

where significantly higher than anticipated counts of influenza-like illness might signal a potential bioterrorism attack.

As an example of such syndromic data, Fricker (2013) describes daily counts of respiratory and gastrointestinal complaints for more than 2 $^1/_2$ years at several hospitals in a large metropolitan area. Table 4.12 presents the respiratory count data from one of these hospitals. There are 980 observations. Fifty observations were missing from the original data set. The missing values were replaced with the last value that was observed on the same day of the week. This type of data imputation is a variation of "Hot Deck Imputation" discussed in Section 1.4.3 and in Fricker (2013). It is also sometimes called last observation (or Value) carried forward (LOCF). For additional discussion see the web site: http://missingdata.lshtm.ac.uk/.

Figure 4.33 is a time series plot of the respiratory syndrome count data in Table 4.12. This plot was constructed using the Graph Builder feature in JMP. This software package overlays a smoothed curve on the data. The curve is fitted using **locally weighted regression**, often called **loess**. This is a variation of kernel regression that uses a weighted average of the data in a local neighborhood around a specific location to determine the value to plot at that location. Loess usually uses either first-order linear regression or a quadratic regression model for the weighted least squares fit. For more information on kernel regression and loess see Montgomery, et al. (2012).

TABLE 4.12 Counts of Respiratory Complaints at a Metropolitan Hospital

Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count
1	17	101	30	201	31	301	12	401	28	501	35	601	26	701	19	801	41	901	29
2	29	102	21	202	23	302	16	402	26	502	27	602	31	702	12	802	50	902	26
3	31	103	32	203	13	303	24	403	28	503	33	603	23	703	17	803	42	903	36
4	34	104	32	204	18	304	21	404	29	504	30	604	24	704	22	804	56	904	31
5	18	105	43	205	36	305	14	405	33	505	30	605	27	705	20	805	36	905	25
6	43	106	25	206	23	306	15	406	36	506	29	606	24	706	22	306	51	906	31
7	34	107	32	207	22	307	23	407	62	507	30	607	31	707	21	807	40	907	32
8	23	108	31	208	23	308	10	403	31	508	22	608	29	708	24	808	29	908	30
9	23	109	33	209	26	309	16	409	30	509	40	609	36	709	16	809	61	909	31
10	39	110	40	210	22	310	11	410	31	510	40	610	31	710	14	810	42	910	29
11	25	111	37	211	21	311	16	411	27	511	41	611	30	711	14	811	56	911	30
12	15	112	34	212	25	312	16	412	35	512	34	612	27	712	30	812	60	912	35
13	29	113	29	213	20	313	12	413	45	513	30	613	27	713	24	813	38	913	24
14	20	114	50	214	18	314	23	414	37	514	33	614	25	714	25	814	52	914	27
15	21	115	27	215	26	315	10	415	23	515	17	615	34	715	17	815	32	915	22
16	22	116	28	216	32	315	15	416	31	516	32	616	33	716	27	816	43	916	33
17	24	117	23	217	41	317	11	417	33	517	40	617	36	717	25	817	54	917	29
18	19	118	27	218	30	318	17	418	27	518	30	618	26	718	14	818	36	913	37
19	28	119	27	219	34	319	13	419	28	519	27	619	20	719	25	819	51	919	29
20	29	120	41	220	38	320	14	420	46	520	30	620	27	720	25	820	57	920	32
21	26	121	29	221	22	321	20	421	39	521	38	621	25	721	26	821	48	921	27
22	22	122	26	222	35	322	10	422	53	522	22	622	36	722	20	822	70	922	22
23	21	123	28	223	36	323	15	423	33	523	27	623	30	723	21	823	48	923	33

(*continued*)

TABLE 4.12 (*Continued*)

Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count
24	29	124	30	224	37	324	14	424	32	524	19	624	39	724	29	824	54	924	29
25	25	125	49	225	27	325	6	425	45	525	19	625	26	725	16	825	36	925	37
26	20	126	43	226	23	326	17	426	21	526	33	626	20	726	24	826	51	926	29
27	20	127	27	227	31	327	17	427	47	527	45	627	27	727	42	827	52	927	20
28	29	128	32	228	39	328	17	423	23	528	34	628	36	728	44	828	48	928	13
29	29	129	13	229	39	329	23	429	39	529	27	629	43	729	34	829	70	929	27
30	32	130	26	230	31	330	9	430	32	530	31	630	46	730	33	830	48	930	23
31	16	131	34	231	43	331	21	431	27	531	19	631	33	731	26	831	57	931	17
32	25	132	27	232	35	332	13	432	29	532	22	632	26	732	29	832	38	932	26
33	20	133	33	233	41	333	13	433	37	533	23	633	33	733	33	833	44	933	23
34	22	134	42	234	24	334	14	434	32	534	13	634	24	734	34	834	34	934	27
35	27	135	29	235	39	335	25	435	28	535	29	635	23	735	42	835	50	935	28
36	32	136	29	236	44	336	15	436	42	536	13	636	51	736	43	836	39	936	21
37	23	137	29	237	35	337	18	437	33	537	20	637	35	737	33	837	65	937	20
38	31	138	28	238	30	338	21	438	36	538	20	638	26	738	31	838	55	938	25
39	22	139	35	239	29	339	18	439	25	539	23	639	32	739	30	839	46	939	30
40	21	140	33	240	13	340	12	440	19	540	17	640	29	740	35	840	57	940	13
41	27	141	3S	241	23	341	10	441	34	541	31	641	24	741	34	841	43	941	19
42	37	142	23	242	19	342	10	442	34	542	21	642	18	742	43	842	50	942	20
43	28	143	28	243	24	343	17	443	33	543	29	643	36	743	21	843	39	943	27
44	41	144	23	244	19	344	12	444	26	544	20	644	15	744	42	844	55	944	14
45	45	145	31	245	27	345	24	445	43	545	21	645	33	745	30	845	38	945	21
46	40	146	29	246	20	345	22	446	31	546	25	646	21	746	29	846	29	946	32
47	32	147	24	247	19	347	14	447	30	547	35	647	25	747	29	347	32	947	18
48	45	148	22	248	28	348	14	448	41	548	24	648	25	748	41	848	27	948	25

49	48	149	30	249	19	349	9	449	15	549	25	649	19	749	35	849	22	949	13
50	51	150	21	250	29	350	19	450	23	550	23	650	23	750	29	850	23	950	25
51	51	151	24	251	24	351	15	451	25	551	27	651	18	751	37	851	25	951	19
52	21	152	21	252	33	352	9	452	27	552	35	652	26	752	31	852	19	952	27
53	43	153	30	253	20	353	18	453	40	553	36	653	27	753	24	853	29	953	27
54	42	154	25	254	29	354	17	454	40	554	33	654	11	754	47	854	34	954	18
55	56	155	17	255	17	355	15	455	34	555	27	655	20	755	3	855	27	955	25
55	51	155	22	255	19	355	21	455	42	555	33	656	13	755	34	355	30	956	26
57	51	157	18	257	23	357	22	457	12	557	25	657	20	757	35	857	30	957	39
58	60	158	19	258	26	358	17	458	24	558	32	658	23	758	39	858	24	958	59
59	35	159	20	259	25	359	21	459	20	559	23	659	19	759	29	859	33	959	34
60	43	160	22	260	32	360	26	460	26	560	42	660	21	760	41	860	29	960	34
61	42	161	39	261	21	361	23	461	46	561	25	661	21	761	36	861	36	961	24
62	55	162	35	262	15	362	20	462	35	562	33	662	29	762	50	862	29	962	25
63	46	163	29	263	20	363	28	463	46	563	19	663	18	763	33	863	27	963	40
64	49	164	24	264	19	364	34	464	33	564	40	664	25	764	38	864	32	964	19
65	40	165	22	265	13	365	23	465	27	565	35	665	24	765	40	865	30	965	35
66	33	166	26	266	25	366	20	466	35	566	36	666	19	766	41	866	23	966	34
67	45	167	27	267	19	367	37	467	33	567	33	667	15	767	34	867	25	967	33
68	37	168	28	268	9	368	22	468	29	568	25	668	23	768	42	868	23	968	29
69	44	169	36	269	20	369	32	469	45	569	33	669	14	769	40	869	29	969	23
70	50	170	31	270	20	370	41	470	18	570	33	670	16	770	50	870	26	970	29
71	37	171	31	271	21	371	35	471	21	571	27	671	16	771	30	871	29	971	25
72	36	172	34	272	21	372	41	472	35	572	33	672	22	772	34	872	22	972	19
73	43	173	19	273	20	373	43	473	39	573	33	673	13	773	28	873	16	973	34
74	49	174	37	274	13	374	33	474	40	574	39	674	19	774	21	874	25	974	37
75	4C	175	39	275	25	375	32	475	33	575	30	675	23	775	24	875	26	975	34

(continued)

TABLE 4.12 ***(Continued)***

Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count	Day	Count
76	65	176	32	276	29	376	28	476	35	576	33	676	14	776	37	876	25	976	29
77	49	177	36	277	16	377	42	477	20	577	26	677	16	777	44	877	24	977	27
78	49	178	42	278	18	378	27	478	36	578	26	678	10	778	39	878	29	978	18
79	34	179	31	279	32	379	25	479	34	579	23	679	14	779	37	879	34	979	26
80	33	180	28	280	32	380	32	480	35	580	24	680	13	780	35	880	35	980	28
81	29	181	35	281	19	381	27	481	36	581	32	681	15	781	32	881	29		
82	32	182	36	282	24	382	35	482	29	582	24	682	15	782	41	882	39		
83	57	183	35	283	18	383	26	483	19	583	32	683	11	783	41	883	31		
84	43	184	32	284	20	384	32	484	36	584	41	634	11	784	51	884	26		
85	40	185	26	285	20	385	42	485	35	585	26	685	18	785	43	885	24		
86	46	186	29	286	20	386	38	486	31	586	28	686	16	786	35	886	31		
87	33	187	25	287	24	387	36	487	23	587	25	687	18	787	33	887	24		
88	30	188	23	288	15	388	26	488	31	588	29	688	15	788	33	888	29		
89	41	189	29	289	22	389	26	489	29	589	40	689	16	789	31	889	26		
90	38	190	29	290	16	390	24	490	44	590	34	690	11	790	43	890	45		
91	29	191	26	291	14	391	30	491	42	591	41	691	11	791	45	891	36		
92	41	192	18	292	17	392	32	492	31	592	37	692	23	792	43	892	29		
93	28	193	19	293	15	393	14	493	31	593	36	693	20	793	42	893	22		
94	47	194	17	294	8	394	27	494	24	594	26	694	18	794	36	894	31		
95	42	195	22	295	23	395	26	495	30	595	42	695	24	795	34	895	38		
96	34	196	25	296	17	396	25	496	26	596	40	696	14	796	30	896	36		
97	40	197	33	297	13	397	23	497	26	597	34	697	22	797	46	897	33		
98	35	198	10	298	15	398	27	498	39	598	41	698	16	798	54	898	34		
99	40	199	25	299	15	399	36	499	35	599	37	699	26	799	52	899	34		
100	24	200	25	300	13	400	40	500	34	600	36	700	17	800	39	900	25		

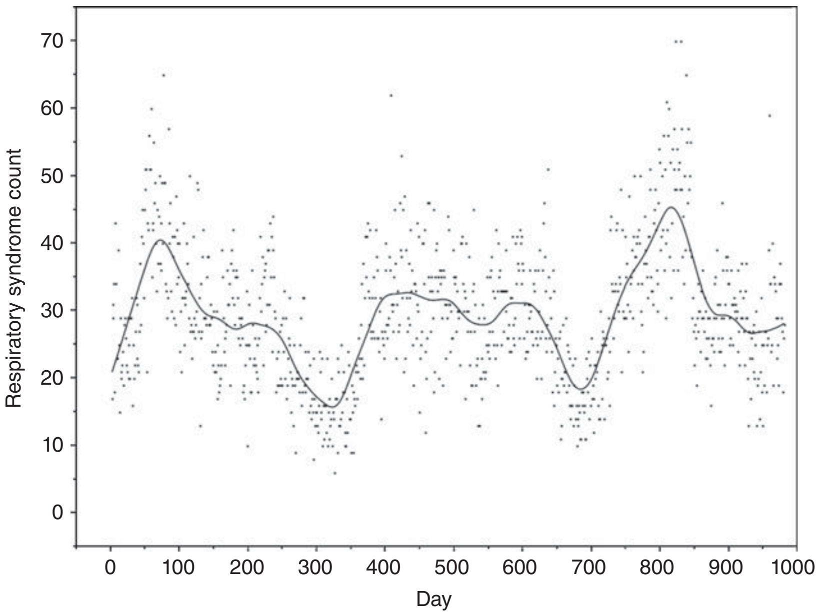


FIGURE 4.33 Time series plot of daily respiratory syndrome count, with kernel-smoothed fitted line. ($\alpha = 0.1$).

Over the 2 $^1/_2$ year period, the daily counts of the respiratory syndrome appear to follow a weak seasonal pattern, with the highest peak in November–December (late fall), a secondary peak in March–April, and then decreasing to the lowest counts in June–August (summer). The amplitude, or range within a year, seems to vary, but counts do not appear to be increasing or decreasing over time.

Not immediately evident from the time series plots is a potential day effect. The box plots of the residuals from the loess smoothed line in Figure 4.33 are plotted in Figure 4.34 versus day of the week. These plots exhibit variation that indicates slightly higher-than-expected counts on Monday and slightly lower-than-expected counts on Thursday, Friday, and Saturday.

The exponential smoothing procedure in JMP was applied to the respiratory syndrome data. The results of first-order or simple exponential smoothing are summarized in Table 4.13 and Figure 4.35, which plots only the last 100 observations along with the smoothed values. JMP reported the value of the smoothing constant that produced the minimum value of the error sum of squares as $\lambda = 0.21$. This value also minimizes the AIC and BIC criteria, and results in the smallest values of the mean absolute

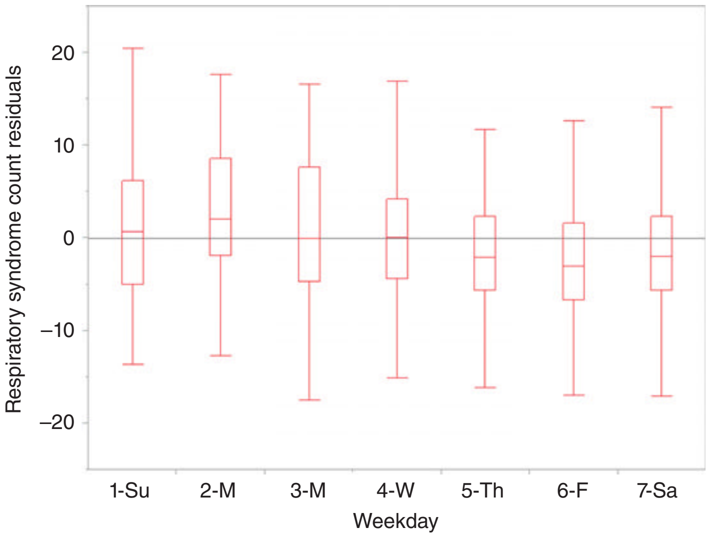


FIGURE 4.34 Box plots of residuals from the kernel-smoothed line fit to daily respiratory syndrome count.

prediction error and the mean absolute, although there is very little difference between the optimal value of $\lambda = 0.21$ and the values $\lambda = 0.1$ and $\lambda = 0.4$.

The results of using second-order exponential smoothing are summarized in Table 4.14 and illustrated graphically for the last 100 observations in Figure 4.36. There is not a lot of difference between the two procedures, although the optimal first-order smoother does perform slightly better and the larger smoothing parameters in the double smoother perform more poorly.

Single and double exponential smoothing do not account for the apparent mild seasonality observed in the original time series plot of the data.

TABLE 4.13 First-Order Simple Exponential Smoothing Applied to the Respiratory Data

Model	Variance	AIC	BIC	MAPE	MAE
First-Order Exponential (min SSE, $\lambda = 0.21$)	52.66	6660.81	6665.70	21.43	5.67
First-Order Exponential ($\lambda = 0.1$)	55.65	6714.67	6714.67	22.23	5.85
First-Order Exponential ($\lambda = 0.4$)	55.21	6705.63	6705.63	21.87	5.82

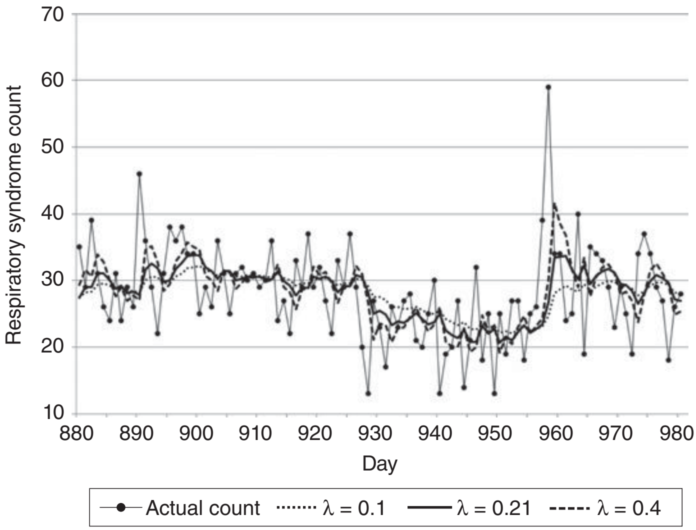


FIGURE 4.35 Respiratory syndrome counts using first-order exponential smoothing with $\lambda = 0.1$, $\lambda = 0.21$ (min SSE), and $\lambda = 0.4$.

We used JMP to fit Winters' additive seasonal model to the respiratory syndrome count data. Because the seasonal patterns are not strong, we investigated seasons of length 3, 7, and 12 periods. The results are summarized in Table 4.15 and illustrated graphically for the last 100 observations in Figure 4.37. The 7-period season works best, probably reflecting the daily seasonal pattern that we observed in Figure 4.34. This is also the best smoother of all the techniques that were investigated. The values of $\lambda = 0$ for the trend and seasonal components in this model are an indication that there is not a significant linear trend in the data and that the seasonal pattern is relatively stable over the period of available data.

TABLE 4.14 Second-Order Simple Exponential Smoothing Applied to the Respiratory Data

Model	Variance	AIC	BIC	MAPE	MAE
Second-Order Exponential (min SSE, $\lambda = 0.10$)	54.37	6690.98	6695.86	21.71	5.78
Second-Order Exponential ($\lambda = 0.2$)	58.22	6754.37	6754.37	22.44	5.98
Second-Order Exponential ($\lambda = 0.4$)	74.46	6992.64	6992.64	25.10	6.74

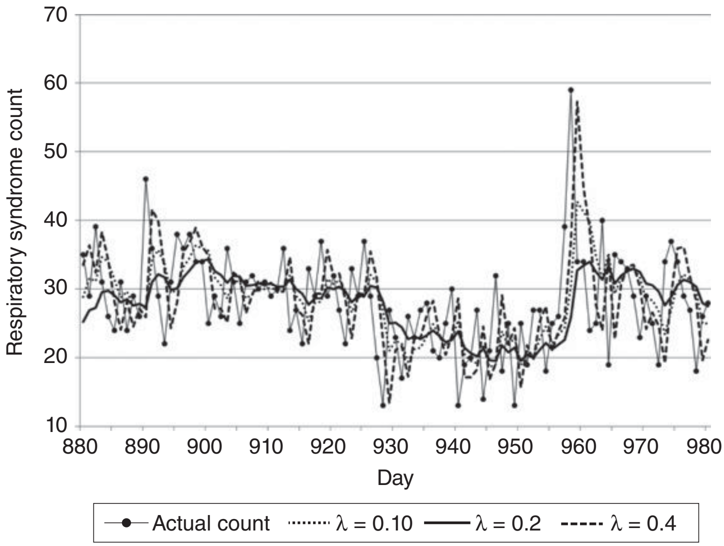


FIGURE 4.36 Respiratory syndrome counts using second-order exponential smoothing with $\lambda = 0.10$ (min SSE), $\lambda = 0.2$, and $\lambda = 0.4$.

TABLE 4.15 Winters' Additive Seasonal Exponential Smoothing Applied to the Respiratory Data

Model	Variance	AIC	BIC	MAPE	MAE
$S = 3$					
Winters Additive (min SSE, $\lambda1 = 0.21$, $\lambda2 = 0$, $\lambda3 = 0$)	52.75	6662.75	6677.40	21.70	5.72
Winters Additive ($\lambda1 = 0.2$, $\lambda2 = 0.1$, $\lambda3 = 0.1$)	57.56	6731.59	6731.59	22.38	5.94
$S = 7$					
Winters Additive (min SSE, $\lambda1 = 0.22$, $\lambda2 = 0$, $\lambda3 = 0$)	49.77	6593.83	6608.47	21.10	5.56
Winters Additive ($\lambda1 = 0.2$, $\lambda2 = 0.1$, $\lambda3 = 0.1$)	54.27	6652.57	6652.57	21.47	5.70
$S = 12$					
Winters Additive (min SSE, $\lambda1 = 0.21$, $\lambda2 = 0$, $\lambda3 = 0$)	52.74	6635.58	6650.21	22.13	5.84
Winters Additive ($\lambda1 = 0.2$, $\lambda2 = 0.1$, $\lambda3 = 0.1$)	58.76	6703.79	6703.79	22.77	6.08

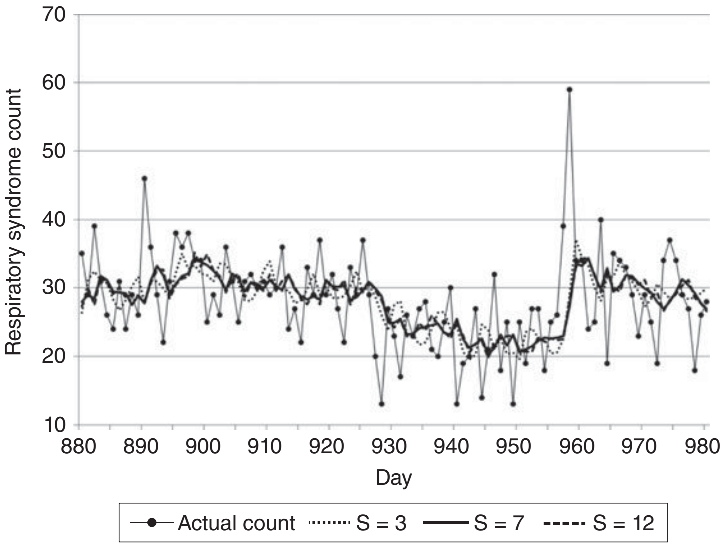


FIGURE 4.37 Respiratory syndrome counts using winters' additive seasonal exponential smoothing with $S = 3$, $S = 7$, and $S = 12$, and smoothing parameters that minimize SSE.

4.9 EXPONENTIAL SMOOTHERS AND ARIMA MODELS

The first-order exponential smoother presented in Section 4.2 is a very effective model in forecasting. The discount factor, λ, makes this smoother fairly flexible in handling time series data with various characteristics. The first-order exponential smoother is particularly good in forecasting time series data with certain specific characteristics.

Recall that the first-order exponential smoother is given as

$$\tilde{y}_T = \lambda y_T + (1 - \lambda)\tilde{y}_{T-1} \tag{4.58}$$

and the forecast error is defined as

$$e_T = y_T - \hat{y}_{T-1}. \tag{4.59}$$

Similarly, we have

$$e_{T-1} = y_{T-1} - \hat{y}_{T-2}. \tag{4.60}$$

By multiplying Eq. (4.60) by $(1 - \lambda)$ and subtracting it from Eq. (4.59), we obtain

$$\begin{aligned} e_T - (1-\lambda)e_{T-1} &= (y_T - \hat{y}_{T-1}) - (1-\lambda)(y_{T-1} - \hat{y}_{T-2}) \\ &= y_T - y_{T-1} - \hat{y}_{T-1} + \underbrace{\lambda y_{T-1} + (1-\lambda)\hat{y}_{T-2}}_{=\hat{y}_{T-1}} \\ &= y_T - y_{T-1} - \hat{y}_{T-1} + \hat{y}_{T-1} \\ &= y_T - y_{T-1}. \end{aligned} \tag{4.61}$$

We can rewrite Eq. (4.61) as

$$y_T - y_{T-1} = e_T - \theta e_{T-1}, \tag{4.62}$$

where $\theta = 1 - \lambda$. Recall from Chapter 2 the **backshift** *operator, B*, defined as $B(y_t) = y_{t-1}$. Thus Eq. (4.62) becomes

$$(1 - B)y_T = (1 - \theta B)e_T. \tag{4.63}$$

We will see in Chapter 5 that the model in Eq. (4.63) is called the **integrated moving average** model denoted as IMA(1,1), for the backshift operator is used only once on y_T and only once on the error. It can be shown that if the process exhibits the dynamics defined in Eq. (4.63), that is an IMA(1,1) process, the first-order exponential smoother provides minimum mean squared error (MMSE) forecasts (see Muth (1960), Box and Luceno (1997), and Box, Jenkins, and Reinsel (1994)). For more discussion of the equivalence between exponential smoothing techniques and the ARIMA models, see Abraham and Ledolter (1983), Cogger (1974), Goodman (1974), Pandit and Wu (1974), and McKenzie (1984).

4.10 R COMMANDS FOR CHAPTER 4

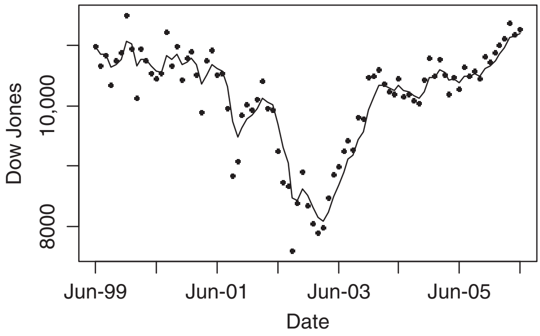
Example 4.1 The Dow Jones index data are in the second column of the array called dji.data in which the first column is the month of the year. We can use the following simple function to obtain the first-order exponential smoothing

```
firstsmooth<-function(y,lambda,start=y[1]){
      ytilde<-y
      ytilde[1]<-lambda*y[1]+(1-lambda)*start
      for (i in 2:length(y)){
            ytilde[i]<-lambda*y[i]+(1-lambda)*ytilde[i-1]
      }
ytilde
}
```

Note that this function uses the first observation as the starting value by default. One can change this by providing a specific start value when calling the function.

We can then obtain the smoothed version of the data for a specified lambda value and plot the fitted value as the following:

```
dji.smooth1<-firstsmooth(y=dji.data[,2],lambda=0.4)
plot(dji.data[,2],type="p", pch=16,cex=.5,xlab='Date',ylab='Dow
    Jones',xaxt='n')
axis(1, seq(1,85,12), dji.data[seq(1,85,12),1])
lines(dji.smooth1)
```



For the first-order exponential smoothing, measures of accuracy such as MAPE, MAD, and MSD can be obtained from the following function:

```
measacc.fs<- function(y,lambda){
            out<- firstsmooth(y,lambda)
            T<-length(y)
            #Smoothed version of the original is the one step
              ahead prediction
            #Hence the predictions (forecasts) are given as
            pred<-c(y[1],out[1:(T-1)])
            prederr<- y-pred
            SSE<-sum(prederr^2)
            MAPE<-100*sum(abs(prederr/y))/T
            MAD<-sum(abs(prederr))/T
            MSD<-sum(prederr^2)/T
            ret1<-c(SSE,MAPE,MAD,MSD)
            names(ret1)<-c("SSE","MAPE","MAD","MSD")
            return(ret1)
}

measacc.fs(dji.data[,2],0.4)
         SSE          MAPE            MAD            MSD
1.665968e+07 3.461342e+00 3.356325e+02 1.959962e+05
```

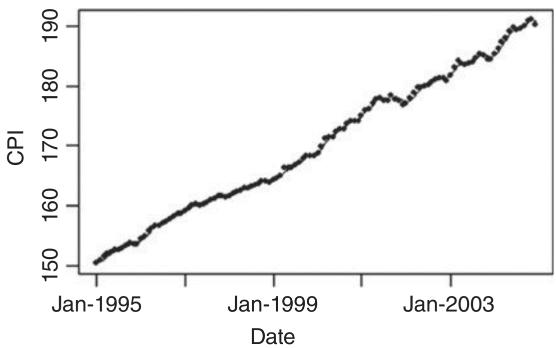
Note that alternatively we could use the Holt–Winters function from the stats package. The function requires three parameters (alpha, beta, and gamma) to be defined. Providing a specific value for alpha and setting beta and gamma to "FALSE" give the first-order exponential as the following

```
dji1.fit<-HoltWinters(dji.data[,2],alpha=.4, beta=FALSE, gamma=FALSE)
```

Beta corresponds to the second-order smoothing (or the trend term) and gamma is for the seasonal effect.

Example 4.2 The US CPI data are in the second column of the array called cpi.data in which the first column is the month of the year. For this case we use the firstsmooth function twice to obtain the double exponential smoothing as

```
cpi.smooth1<-firstsmooth(y=cpi.data[,2],lambda=0.3)
cpi.smooth2<-firstsmooth(y=cpi.smooth1,lambda=0.3)
cpi.hat<-2*cpi.smooth1-cpi.smooth2 #Equation 4.23
plot(cpi.data[,2],type="p", pch=16,cex=.5,xlab='Date',ylab='CPI',
  xaxt='n')
axis(1, seq(1,120,24), cpi.data[seq(1,120,24),1])
lines(cpi.hat)
```

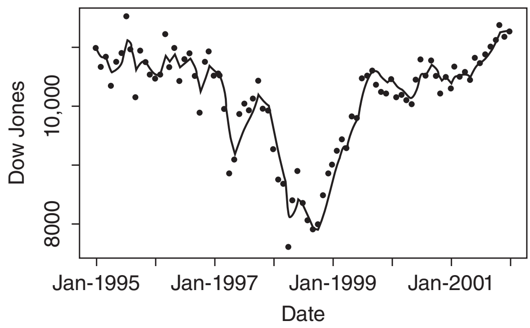


Note that the fitted values are obtained using Eq. (4.23). Also the corresponding command using Holt–Winters function is

```
HoltWinters(cpi.data[,2],alpha=0.3, beta=0.3, gamma=FALSE)
```

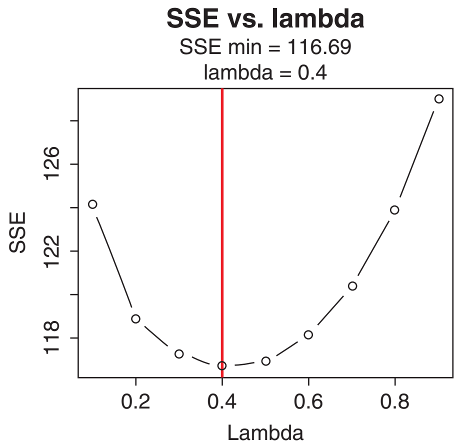
Example 4.3 In this example we use the firstsmooth function twice for the Dow Jones Index data to obtain the double exponential smoothing as in the previous example.

```
dji.smooth1<-firstsmooth(y=dji.data[,2],lambda=0.3)
dji.smooth2<-firstsmooth(y=dji.smooth1,lambda=0.3)
dji.hat<-2*dji.smooth1-dji.smooth2 #Equation 4.23
plot(dji.data[,2],type="p", pch=16,cex=.5,xlab='Date',ylab='Dow
 Jones',xaxt='n')
axis(1, seq(1,85,12), cpi.data[seq(1,85,12),1])
lines(dji.hat)
```



Example 4.4 The average speed data are in the second column of the array called speed.data in which the first column is the index for the week. To find the "best" smoothing constant, we will use the firstsmooth function for various lambda values and obtain the sum of squared one-step-ahead prediction error (SS_E) for each. The lambda value that minimizes the sum of squared prediction errors is deemed the "best" lambda. The obvious option is to apply firstsmooth function in a for loop to obtain SS_E for various lambda values. Even though in this case this may not be an issue, in many cases for loops can slow down the computations in R and are to be avoided if possible. We will do that using sapply function.

```
lambda.vec<-seq(0.1, 0.9, 0.1)
sse.speed<-function(sc){measacc.fs(speed.data[1:78,2],sc)[1]}
sse.vec<-sapply(lambda.vec, sse.speed)
opt.lambda<-lambda.vec[sse.vec == min(sse.vec)]
plot(lambda.vec, sse.vec, type="b", main = "SSE vs. lambda\n",
  xlab='lambda\n',ylab='SSE')
abline(v=opt.lambda, col = 'red')
mtext(text = paste("SSE min = ", round(min(sse.vec),2), "\n lambda
  = ", opt.lambda))
```



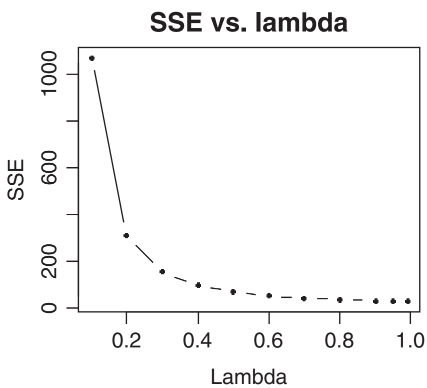
Note that we can also use Holt–Winters function to find the "best" value for the smoothing constant by not specifying the appropriate parameter as the following:

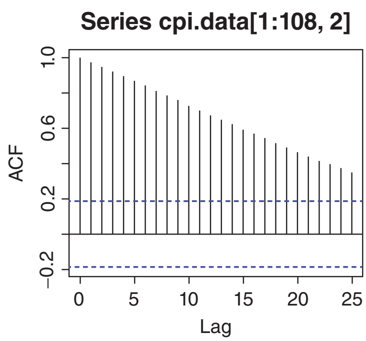
```
HoltWinters(speed.data[,2], beta=FALSE, gamma=FALSE)
```

Example 4.5 We will first try to find the best lambda for the CPI data using first-order exponential smoothing. We will also plot ACF of the data.

Note that we will use the data up to December 2003.

```
lambda.vec<-c(seq(0.1, 0.9, 0.1), .95, .99)
sse.cpi<-function(sc){measacc.fs(cpi.data[1:108,2],sc)[1]}
sse.vec<-sapply(lambda.vec, sse.cpi)
opt.lambda<-lambda.vec[sse.vec == min(sse.vec)]
plot(lambda.vec, sse.vec, type="b", main = "SSE vs. lambda\n",
  xlab='lambda\n',ylab='SSE', pch=16,cex=.5)
acf(cpi.data[1:108,2],lag.max=25)
```

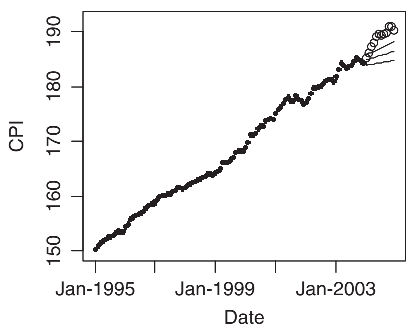




We now use the second-order exponential smoothing with lambda of 0.3. We calculate the forecasts using Eq. (4.31) for the two options suggested in the Example 4.5.

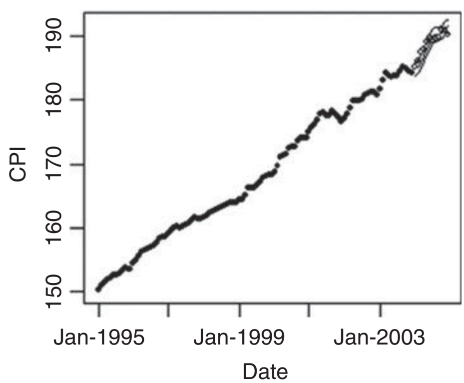
Option 1: On December 2003, make the forecasts for the entire 2004 (1- to 12-step-ahead forecasts).

```
lcpi<-0.3
cpi.smooth1<-firstsmooth(y=cpi.data[1:108,2],lambda=lcpi)
cpi.smooth2<-firstsmooth(y=cpi.smooth1,lambda=lcpi)
cpi.hat<-2*cpi.smooth1-cpi.smooth2
tau<-1:12
T<-length(cpi.smooth1)
cpi.forecast<-(2+tau*(lcpi/(1-lcpi)))*cpi.smooth1[T]-(1+tau*(lcpi/
  (1-lcpi)))*cpi.smooth2[T]
ctau<-sqrt(1+(lcpi/((2-lcpi)^3))*(10-14*lcpi+5*(lcpi^2)+2*tau*lcpi
  *(4-3*lcpi)+2*(tau^2)*(lcpi^2)))
alpha.lev<-.05
sig.est<- sqrt(var(cpi.data[2:108,2]- cpi.hat[1:107]))
cl<-qnorm(1-alpha.lev/2)*(ctau/ctau[1])*sig.est
plot(cpi.data[1:108,2],type="p", pch=16,cex=.5,xlab='Date',
  ylab='CPI',xaxt='n',xlim=c(1,120),ylim=c(150,192))
axis(1, seq(1,120,24), cpi.data[seq(1,120,24),1])
points(109:120,cpi.data[109:120,2])
lines(109:120,cpi.forecast)
lines(109:120,cpi.forecast+cl)
lines(109:120,cpi.forecast-cl)
```



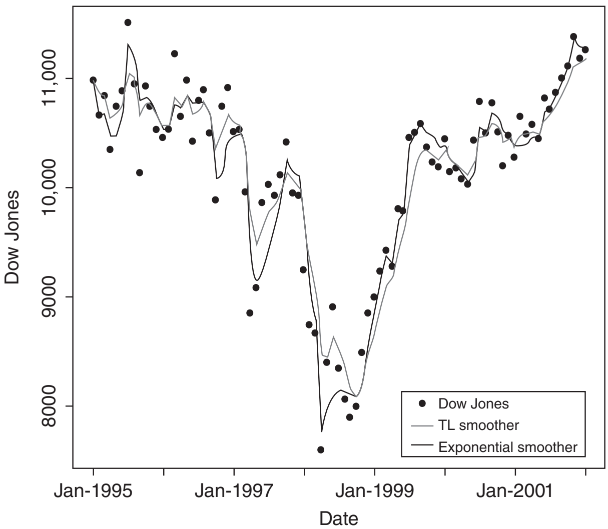
Option 2: On December 2003, make the forecast for January 2004. Then when January 2004 data are available, make the forecast for February 2004 (only one-step-ahead forecasts).

```
lcpi<-0.3
T<-108
tau<-12
alpha.lev<-.05
cpi.forecast<-rep(0,tau)
cl<-rep(0,tau)
cpi.smooth1<-rep(0,T+tau)
cpi.smooth2<-rep(0,T+tau)

for (i in 1:tau) {
      cpi.smooth1[1:(T+i-1)]<-firstsmooth(y=cpi.data[1:(T+i-1),2],
        lambda=lcpi)
      cpi.smooth2[1:(T+i-1)]<-firstsmooth(y=cpi.smooth1[1:(T+i-1)],
        lambda=lcpi)
      cpi.forecast[i]<-(2+(lcpi/(1-lcpi)))*cpi.smooth1[T+i-1]-
        (1+(lcpi/(1-lcpi)))*cpi.smooth2[T+i-1]
      cpi.hat<-2*cpi.smooth1[1:(T+i-1)]-cpi.smooth2[1:(T+i-1)]
      sig.est<- sqrt(var(cpi.data[2:(T+i-1),2]- cpi.hat[1:(T+i-2)]))
      cl[i]<-qnorm(1-alpha.lev/2)*sig.est
      }

plot(cpi.data[1:T,2],type="p", pch=16,cex=.5,xlab='Date',ylab='CPI',
  xaxt='n',xlim=c(1,T+tau),ylim=c(150,192))
axis(1, seq(1,T+tau,24), cpi.data[seq(1,T+tau,24),1])
points((T+1):(T+tau),cpi.data[(T+1):(T+tau),2],cex=.5)
lines((T+1):(T+tau),cpi.forecast)
lines((T+1):(T+tau),cpi.forecast+cl)
lines((T+1):(T+tau),cpi.forecast-cl)
```



Example 4.6 The function for the Trigg–Leach smoother is given as:

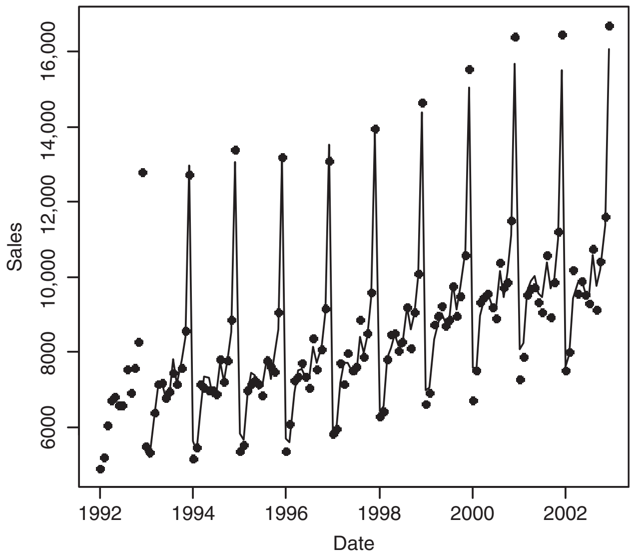
```
tlsmooth<-function(y,gamma,y.tilde.start=y[1],lambda.start=1){
      T<-length(y)

#Initialize the vectors
      Qt<-vector()
      Dt<-vector()
      y.tilde<-vector()
      lambda<-vector()
      err<-vector()
#Set the starting values for the vectors
      lambda[1]=lambda.start
      y.tilde[1]=y.tilde.start
      Qt[1]<-0
      Dt[1]<-0
      err[1]<-0

      for (i in 2:T){
            err[i]<-y[i]-y.tilde[i-1]
            Qt[i]<-gamma*err[i]+(1-gamma)*Qt[i-1]
            Dt[i]<-gamma*abs(err[i])+(1-gamma)*Dt[i-1]
            lambda[i]<-abs(Qt[i]/Dt[i])
            y.tilde[i]=lambda[i]*y[i] + (1-lambda[i])*y.tilde[i-1]
      }
return(cbind(y.tilde,lambda,err,Qt,Dt))
}

#Obtain the TL smoother for Dow Jones Index
out.tl.dji<-tlsmooth(dji.data[,2],0.3)
```



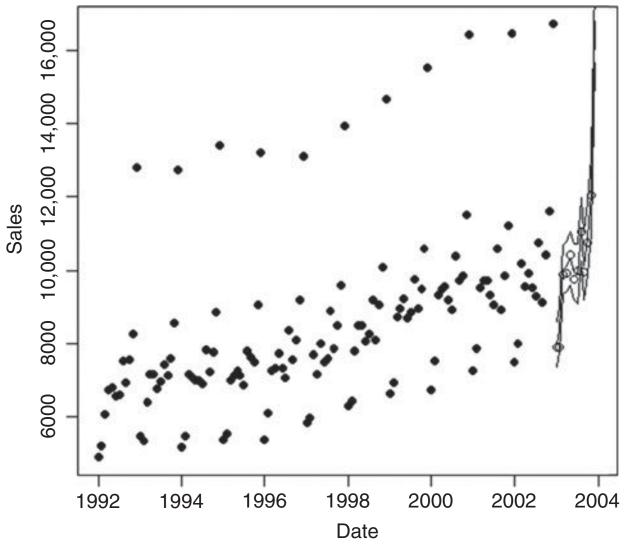
```
#Obtain the exponential smoother for Dow Jones Index
dji.smooth1<-firstsmooth(y=dji.data[,2],lambda=0.4)

#Plot the data together with TL and exponential smoother for
  comparison
plot(dji.data[,2],type="p", pch=16,cex=.5,xlab='Date',ylab='Dow
  Jones',xaxt='n')
axis(1, seq(1,85,12), cpi.data[seq(1,85,12),1])
lines(out.tl.dji[,1])
lines(dji.smooth1,col="grey40")
legend(60,8000,c("Dow Jones","TL Smoother","Exponential Smoother"),
  pch=c(16, NA, NA),lwd=c(NA,.5,.5),cex=.55,col=c("black",
  "black","grey40"))
```

Example 4.7 The clothing sales data are in the second column of the array called closales.data in which the first column is the month of the year. We will use the data up to December 2002 to fit the model and make forecasts for the coming year (2003). We will use Holt–Winters function given in stats package. The model is additive seasonal model with all parameters equal to 0.2.

```
dat.ts = ts(closales.data[,2], start = c(1992,1), freq = 12)
y1<-closales.data[1:132,]
# convert data to ts object
y1.ts<-ts(y1[,2], start = c(1992,1), freq = 12)
clo.hw1<-HoltWinters(y1.ts,alpha=0.2,beta=0.2,gamma=0.2,seasonal
  ="additive")
plot(y1.ts,type="p", pch=16,cex=.5,xlab='Date',ylab='Sales')
lines(clo.hw1$fitted[,1])
```

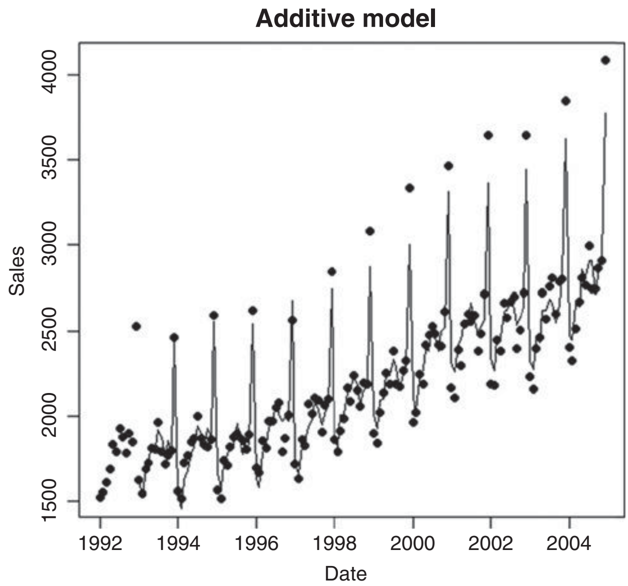


```
#Forecast the the sales for 2003
y2<-closales.data[133:144,]
y2.ts<-ts(y2[,2],start=c(2003,1),freq=12)

y2.forecast<-predict(clo.hw1, n.ahead=12, prediction.interval
  = TRUE)
plot(y1.ts,type="p", pch=16,cex=.5,xlab='Date',ylab='Sales',
  xlim=c(1992,2004))
points(y2.ts)
lines(y2.forecast[,1])
lines(y2.forecast[,2])
lines(y2.forecast[,3])
```

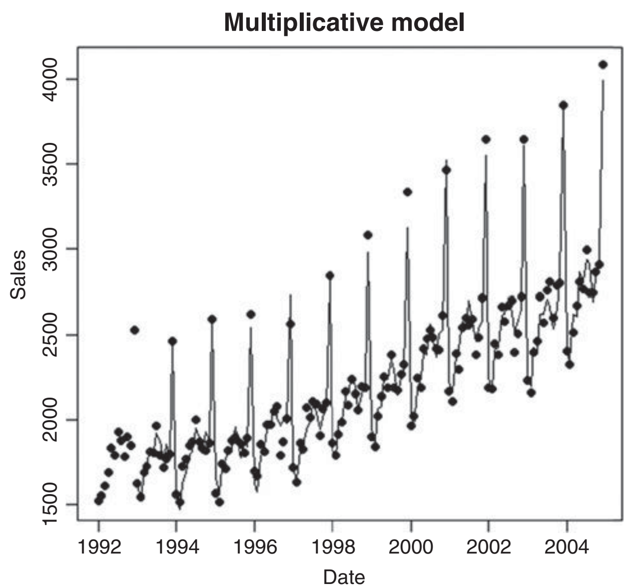


Example 4.8 The liquor store sales data are in the second column of the array called liqsales.data in which the first column is the month of the year. We will first fit additive and multiplicative seasonal models to the entire data to see the difference in the fits. Then we will use the data up to December 2003 to fit the multiplicative model and make forecasts for the coming year (2004). We will once again use Holt–Winters function given in stats package. In all cases we set all parameters to 0.2.

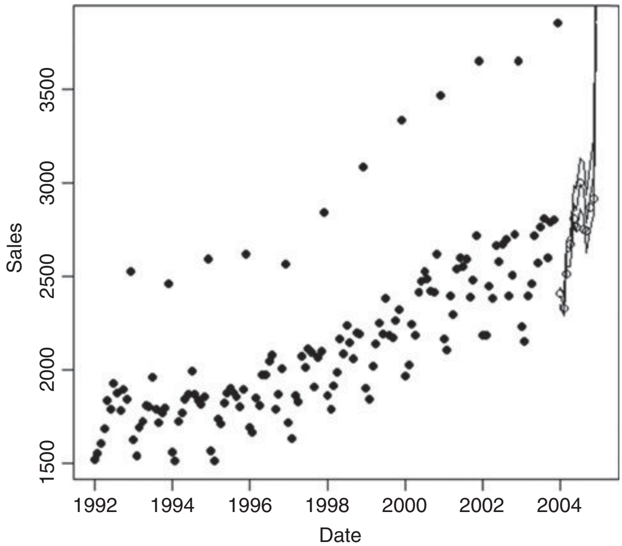
```
y.ts<- ts(liqsales.data[,2], start = c(1992,1), freq = 12)

liq.hw.add<-HoltWinters(y.ts,alpha=0.2,beta=0.2,gamma=0.2,
  seasonal="additive")
plot(y.ts,type="p", pch=16,cex=.5,xlab='Date',ylab='Sales',
  main="Additive Model")
lines(liq.hw.add$fitted[,1])
```




```
liq.hw.mult<-HoltWinters(y.ts,alpha=0.2,beta=0.2,gamma=0.2,
  seasonal="multiplicative")
plot(y.ts,type="p", pch=16,cex=.5,xlab='Date',ylab='Sales',
  main="Multiplicative Model")
lines(liq.hw.mult$fitted[,1])
```



```
y1<-liqsales.data[1:144,]
y1.ts<-ts(y1[,2], start = c(1992,1), freq = 12)
liq.hw1<-HoltWinters(y1.ts,alpha=0.2,beta=0.2,gamma=0.2,
  seasonal="multiplicative")
y2<-liqsales.data[145:156,]
y2.ts<-ts(y2[,2],start=c(2004,1),freq=12)

y2.forecast<-predict(liq.hw1, n.ahead=12, prediction.interval =
  TRUE)
plot(y1.ts,type="p", pch=16,cex=.5,xlab='Date',ylab='Sales',
  xlim=c(1992,2005))
points(y2.ts)
lines(y2.forecast[,1])
lines(y2.forecast[,2])
lines(y2.forecast[,3])
```



EXERCISES

4.1 Consider the time series data shown in Table E4.1.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the first 40 time periods of this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

TABLE E4.1 Data for Exercise 4.1

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	48.7	11	49.1	21	45.3	31	50.8	41	47.9
2	45.8	12	46.7	22	43.3	32	46.4	42	49.5
3	46.4	13	47.8	23	44.6	33	52.3	43	44.0
4	46.2	14	45.8	24	47.1	34	50.5	44	53.8
5	44.0	15	45.5	25	53.4	35	53.4	45	52.5
6	53.8	16	49.2	26	44.9	36	53.9	46	52.0
7	47.6	17	54.8	27	50.5	37	52.3	47	50.6
8	47.0	18	44.7	28	48.1	38	53.0	48	48.7
9	47.6	19	51.1	29	45.4	39	48.6	49	51.4
10	51.1	20	47.3	30	51.6	40	52.4	50	47.7

4.2 Reconsider the time series data shown in Table E4.1.

a. Use simple exponential smoothing with the optimum value of λ to smooth the first 40 time periods of this data (you can find the optimum value from Minitab). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.1.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors. Compare these forecast errors with those from Exercise 4.1. How much has using the optimum value of the smoothing constant improved the forecasts?

4.3 Find the sample ACF for the time series in Table E4.1. Does this give you any insight about the optimum value of the smoothing constant that you found in Exercise 4.2?

4.4 Consider the time series data shown in Table E4.2.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the first 40 time periods of this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

4.5 Reconsider the time series data shown in Table E4.2.

a. Use simple exponential smoothing with the optimum value of λ to smooth the first 40 time periods of this data (you can find the optimum value from Minitab). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.4.

TABLE E4.2 Data for Exercise 4.4

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	43.1	11	41.8	21	47.7	31	52.9	41	48.3
2	43.7	12	50.7	22	51.1	32	47.3	42	45.0
3	45.3	13	55.8	23	67.1	33	50.0	43	55.2
4	47.3	14	48.7	24	47.2	34	56.7	44	63.7
5	50.6	15	48.2	25	50.4	35	42.3	45	64.4
6	54.0	16	46.9	26	44.2	36	52.0	46	66.8
7	46.2	17	47.4	27	52.0	37	48.6	47	63.3
8	49.3	18	49.2	28	35.5	38	51.5	48	60.0
9	53.9	19	50.9	29	48.4	39	49.5	49	60.9
10	42.5	20	55.3	30	55.4	40	51.4	50	56.1

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors. Compare these forecast errors with those from Exercise 4.4. How much has using the optimum value of the smoothing constant improved the forecasts?

4.6 Find the sample ACF for the time series in Table E4.2. Does this give you any insight about the optimum value of the smoothing constant that you found in Exercise 4.5?

4.7 Consider the time series data shown in Table E4.3.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth the first 30 time periods of this data. How well does this smoothing procedure work?

TABLE E4.3 Data for Exercise 4.7

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	275	11	297	21	231	31	255	41	293
2	245	12	235	22	238	32	255	42	284
3	222	13	237	23	251	33	229	43	276
4	169	14	203	24	253	34	286	44	290
5	236	15	238	25	283	35	236	45	250
6	259	16	232	26	283	36	194	46	235
7	268	17	206	27	245	37	228	47	275
8	225	18	295	28	234	38	244	48	350
9	246	19	247	29	273	39	241	49	290
10	263	20	227	30	293	40	284	50	269

c. Make one-step-ahead forecasts of the last 20 observations. Determine the forecast errors.

d. Plot the forecast errors on a control chart for individuals. Use a moving range chart to estimate the standard deviation of the forecast errors in constructing this chart. What conclusions can you draw about the forecasting procedure and the time series?

4.8 The data in Table E4.4 exhibit a linear trend.

a. Verify that there is a trend by plotting the data.

b. Using the first 12 observations, develop an appropriate procedure for forecasting.

c. Forecast the last 12 observations and calculate the forecast errors. Does the forecasting procedure seem to be working satisfactorily?

TABLE E4.4 Data for Exercise 4.8

Period	y_t	Period	y_t
1	315	13	460
2	195	14	395
3	310	15	390
4	316	16	450
5	325	17	458
6	335	18	570
7	318	19	520
8	355	20	400
9	420	21	420
10	410	22	580
11	485	23	475
12	420	24	560

4.9 Reconsider the linear trend data in Table E4.4. Take the first difference of this data and plot the time series of first differences. Has differencing removed the trend? Use exponential smoothing on the first 11 differences. Instead of forecasting the original data, forecast the first differences for the remaining data using exponential smoothing and use these forecasts of the first differences to obtain forecasts for the original data.

4.10 Table B.1 in Appendix B contains data on the market yield on US Treasury Securities at 10-year constant maturity.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the data, excluding the last 20 observations. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 20 observations. Determine the forecast errors.

4.11 Reconsider the US Treasury Securities data shown in Table B.1.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data, excluding the last 20 observations (you can find the optimum value from Minitab). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.10.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors. Compare these forecast errors with those from Exercise 4.10. How much has using the optimum value of the smoothing constant improved the forecasts?

4.12 Table B.2 contains data on pharmaceutical product sales.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

4.13 Reconsider the pharmaceutical sales data shown in Table B.2.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.12.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors. Compare these forecast errors with those from Exercise 4.12. How much has using the optimum value of the smoothing constant improved the forecasts?

c. Construct the sample ACF for these data. Does this give you any insight regarding the optimum value of the smoothing constant?

4.14 Table B.3 contains data on chemical process viscosity.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

4.15 Reconsider the chemical process data shown in Table B.3.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.14.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors. Compare these forecast errors with those from Exercise 4.14.How much has using the optimum value of the smoothing constant improved the forecasts?

c. Construct the sample ACF for these data. Does this give you any insight regarding the optimum value of the smoothing constant?

4.16 Table B.4 contains data on the annual US production of blue and gorgonzola cheeses. This data have a strong trend.

a. Verify that there is a trend by plotting the data.

b. Develop an appropriate exponential smoothing procedure for forecasting.

c. Forecast the last 10 observations and calculate the forecast errors. Does the forecasting procedure seem to be working satisfactorily?

4.17 Reconsider the blue and gorgonzola cheese data in Table B.4 and Exercise 4.16. Take the first difference of this data and plot the time series of first differences. Has differencing removed the trend? Use exponential smoothing on the first differences. Instead of forecasting the original data, develop a procedure for forecasting the first differences and explain how you would use these forecasts of the first differences to obtain forecasts for the original data.

4.18 Table B.5 shows data for US beverage manufacturer product shipments. Develop an appropriate exponential smoothing procedure for forecasting these data.

4.19 Table B.6 contains data on the global mean surface air temperature anomaly.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the data. How well does this smoothing procedure work? Do you think this would be a reliable forecasting procedure?

4.20 Reconsider the global mean surface air temperature anomaly data shown in Table B.6 and used in Exercise 4.19.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.19.

b. Do you think using the optimum value of the smoothing constant would result in improved forecasts from exponential smoothing?

c. Take the first difference of this data and plot the time series of first differences. Use exponential smoothing on the first differences. Instead of forecasting the original data, develop a procedure for forecasting the first differences and explain how you would use these forecasts of the first differences to obtain forecasts for the original global mean surface air temperature anomaly.

4.21 Table B.7 contains daily closing stock prices for the Whole Foods Market.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth the data. How well does this smoothing procedure work? Do you think this would be a reliable forecasting procedure?

4.22 Reconsider the Whole Foods Market data shown in Table B.7 and used in Exercise 4.21.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.21.

b. Do you think that using the optimum value of the smoothing constant would result in improved forecasts from exponential smoothing?

c. Use an exponential smoothing procedure for trends on this data. Is this an apparent improvement over the use of simple exponential smoothing with the optimum smoothing constant?

d. Take the first difference of this data and plot the time series of first differences. Use exponential smoothing on the first differences. Instead of forecasting the original data, develop a procedure for forecasting the first differences and explain how you would use these forecasts of the first differences to obtain forecasts for the stock price.

4.23 Unemployment rate data are given in Table B.8.

a. Make a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the data. How well does this smoothing procedure work? Do you think that simple exponential smoothing should be used to forecast this data?

4.24 Reconsider the unemployment rate data shown in Table B.8 and used in Exercise 4.23.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.23.

b. Do you think that using the optimum value of the smoothing constant would result in improved forecasts from exponential smoothing?

c. Use an exponential smoothing procedure for trends on this data. Is this an apparent improvement over the use of simple exponential smoothing with the optimum smoothing constant?

d. Take the first difference of this data and plot the time series of first differences. Use exponential smoothing on the first differences. Is this a reasonable procedure for forecasting the first differences?

4.25 Table B.9 contains yearly data on the international sunspot numbers.

a. Construct a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth the data. How well does this smoothing procedure work? Do you think that simple exponential smoothing should be used to forecast this data?

4.26 Reconsider the sunspot data shown in Table B.9 and used in Exercise 4.25.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.25.

b. Do you think that using the optimum value of the smoothing constant would result in improved forecasts from exponential smoothing?

c. Use an exponential smoothing procedure for trends on this data. Is this an apparent improvement over the use of simple exponential smoothing with the optimum smoothing constant?

4.27 Table B.10 contains 7 years of monthly data on the number of airline miles flown in the United Kingdom. This is seasonal data.

a. Make a time series plot of the data and verify that it is seasonal.

b. Use Winters' multiplicative method for the first 6 years to develop a forecasting method for this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

4.28 Reconsider the airline mileage data in Table B.10 and used in Exercise 4.27.

a. Use the additive seasonal effects model for the first 6 years to develop a forecasting method for this data. How well does this smoothing procedure work?

b. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

c. Compare these forecasts with those found using Winters' multiplicative method in Exercise 4.27.

4.29 Table B.11 contains 8 years of monthly champagne sales data. This is seasonal data.

a. Make a time series plot of the data and verify that it is seasonal. Why do you think seasonality is present in these data?

b. Use Winters' multiplicative method for the first 7 years to develop a forecasting method for this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

4.30 Reconsider the monthly champagne sales data in Table B.11 and used in Exercise 4.29.

a. Use the additive seasonal effects model for the first 7 years to develop a forecasting method for this data. How well does this smoothing procedure work?

b. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

c. Compare these forecasts with those found using Winters' multiplicative method in Exercise 4.29.

4.31 Montgomery et al. (1990) give 4 years of data on monthly demand for a soft drink. These data are given in Table E4.5.

a. Make a time series plot of the data and verify that it is seasonal. Why do you think seasonality is present in these data?

b. Use Winters' multiplicative method for the first 3 years to develop a forecasting method for this data. How well does this smoothing procedure work?

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

TABLE E4.5 Soft Drink Demand Data

Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	143	13	189	25	359	37	332
2	191	14	326	26	264	38	244
3	195	15	289	27	315	39	320
4	225	16	293	28	362	40	437
5	175	17	279	29	414	41	544
6	389	18	552	30	647	42	830
7	454	19	674	31	836	43	1011
8	618	20	827	32	901	44	1081
9	770	21	1000	33	1104	45	1400
10	564	22	502	34	874	46	1123
11	327	23	512	35	683	47	713
12	235	24	300	36	352	48	487

4.32 Reconsider the soft drink demand data in Table E4.5 and used in Exercise 4.31.

a. Use the additive seasonal effects model for the first 3 years to develop a forecasting method for this data. How well does this smoothing procedure work?

b. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

c. Compare these forecasts with those found using Winters' multiplicative method in Exercise 4.31.

4.33 Table B.12 presents data on the hourly yield from a chemical process and the operating temperature. Consider only the yield data in this exercise.

a. Construct a time series plot of the data.

b. Use simple exponential smoothing with $\lambda = 0.2$ to smooth the data. How well does this smoothing procedure work? Do you think that simple exponential smoothing should be used to forecast this data?

4.34 Reconsider the chemical process yield data shown in Table B.12.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP). How well does this smoothing procedure work? Compare the results with those obtained in Exercise 4.33.

b. How much has using the optimum value of the smoothing constant improved the forecasts?

4.35 Find the sample ACF for the chemical process yield data in Table B.12. Does this give you any insight about the optimum value of the smoothing constant that you found in Exercise 4.34?

4.36 Table B.13 presents data on ice cream and frozen yogurt sales. Develop an appropriate exponential smoothing forecasting procedure for this time series.

4.37 Table B.14 presents the CO_2 readings from Mauna Loa.

a. Use simple exponential smoothing with the optimum value of λ to smooth the data (you can find the optimum value from either Minitab or JMP).

b. Use simple exponential smoothing with $\lambda = 0.1$ to smooth the data. How well does this smoothing procedure work? Compare the results with those obtained using the optimum smoothing constant. How much has using the optimum value of the smoothing constant improved the exponential smoothing procedure?

4.38 Table B.15 presents data on the occurrence of violent crimes. Develop an appropriate exponential smoothing forecasting procedure for this time series.

4.39 Table B.16 presents data on the US. gross domestic product (GDP). Develop an appropriate exponential smoothing forecasting procedure for the GDP time series.

4.40 Total annual energy consumption is shown in Table B.17. Develop an appropriate exponential smoothing forecasting procedure for the energy consumption time series.

4.41 Table B.18 contains data on coal production. Develop an appropriate exponential smoothing forecasting procedure for the coal production time series.

4.42 Table B.19 contains data on the number of children 0−4 years old who drowned in Arizona.

a. Plot the data. What type of forecasting model seems appropriate?

b. Develop a forecasting model for this data?

4.43 Data on tax refunds and population are shown in Table B.20. Develop an appropriate exponential smoothing forecasting procedure for the tax refund time series.

4.44 Table B.21 contains data from the US Energy Information Administration on monthly average price of electricity for the residential sector in Arizona. This data have a strong seasonal component. Use the data from 2001–2010 to develop a multiplicative Winters-type exponential smoothing model for this data. Use this model to simulate one-month-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

4.45 Use the electricity price data in Table B.21 from 2010–2010 and an additive Winters-type exponential smoothing procedure to develop a forecasting model.

a. Use this model to simulate one-month-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

b. Compare the performance of this model with the multiplicative model you developed in Exercise 4.44.

4.46 Table B.22 contains data from the Danish Energy Agency on Danish crude oil production.

a. Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

b. Use first-order exponential smoothing to develop a forecasting model for crude oil production. Plot the smoothed statistic on the same axes with the original data. How well does first-order exponential smoothing seem to work?

c. Use double exponential smoothing to develop a forecasting model for crude oil production. Plot the smoothed statistic on the same axes with the original data. How well does double exponential smoothing seem to work?

d. Compare the two smoothing models from parts b and c. Which approach seems preferable?

4.47 Apply a first difference to the Danish crude oil production data in Table B.22.

a. Plot the data and comment on any features that you observe from the graph. Calculate and plot the sample ACF and variogram. Interpret these graphs.

b. Use first-order exponential smoothing to develop a forecasting model for crude oil production. Plot the smoothed statistic on the same axes with the original data. How well does first-order exponential smoothing seem to work? How does this compare to the first-order exponential smoothing model you developed in Exercise 4.46 for the original (undifferenced) data?

4.48 Table B.23 shows weekly data on positive laboratory test results for influenza. Notice that these data have a number of missing values. In exercise you were asked to develop and implement a scheme to estimate the missing values. This data have a strong seasonal component. Use the data from 1997–2010 to develop a multiplicative Winters-type exponential smoothing model for this data. Use this model to simulate one-week-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

4.49 Repeat Exercise 4.48 using an additive Winters-type model. Compare the performance of the additive and the multiplicative model from Exercise 4.48.

4.50 Data from the Western Regional Climate Center for the monthly mean daily solar radiation (in Langleys) at the Zion Canyon, Utah, station are shown in Table B.24. This data have a strong seasonal component. Use the data from 2003–2012 to develop a multiplicative Winters-type exponential smoothing model for this data. Use this model to simulate one-month-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

4.51 Repeat Exercise 4.50 using an additive Winters-type model. Compare the performance of the additive and the multiplicative model from Exercise 4.50.

4.52 Table B.25 contains data from the National Highway Traffic Safety Administration on motor vehicle fatalities from 1966 to 2012. This data are used by a variety of governmental and industry groups, as well as research organizations.

a. Plot the fatalities data and comment on any features of the data that you see.

b. Develop a forecasting procedure using first-order exponential smoothing. Use the data from 1966–2006 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

c. Develop a forecasting procedure using based on double exponential smoothing. Use the data from 1966–2006 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work in comparison to the method based on first-order exponential smoothing?

4.53 Apply a first difference to the motor vehicle fatalities data in Table B.25.

a. Plot the differenced data and comment on any features of the data that you see.

b. Develop a forecasting procedure for the first differences based on first-order exponential smoothing. Use the data from 1966–2006 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

c. Compare this approach with the two smoothing methods used in Exercise 4.52.

4.54 Appendix Table B.26 contains data on monthly single-family residential new home sales from 1963 through 2014.

a. Plot the home sales data. Comment on the graph.

b. Develop a forecasting procedure using first-order exponential smoothing. Use the data from1963–2000 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

c. Can you explain the unusual changes in sales observed in the data near the end of the graph?

4.55 Appendix Table B.27 contains data on the airline's best on-time arrival and airport performance. The data are given by month from January 1995 through February 2013.

a. Plot the data and comment on any features of the data that you see.

b. Construct the sample ACF and variogram. Comment on these displays.

c. Develop an appropriate exponential smoothing model for these data.

4.56 Data from the US Census Bureau on monthly domestic automobile manufacturing shipments (in millions of dollars) are shown in Table B.28.

a. Plot the data and comment on any features of the data that you see.

b. Construct the sample ACF and variogram. Comment on these displays.

c. Develop an appropriate exponential smoothing model for these data. Note that there is some apparent seasonality in the data. Why does this seasonal behavior occur?

d. Plot the first difference of the data. Now compute the sample ACF and variogram for the differenced data. What impact has differencing had? Is there still some apparent seasonality in the differenced data?

4.57 Suppose that simple exponential smoothing is being used to forecast a process. At the start of period $t*$, the mean of the process shifts to a new level $\mu + \delta$. The mean remains at this new level for subsequent time periods. Show that the expected value of the exponentially smoothed statistic is

$$E(\hat{y}_t) = \begin{cases} \mu, & T \leq t^* \\ \mu + \delta - \delta(1 - \lambda)^{T-t^*+1}, & T \geq t^* \end{cases}.$$

4.58 Using the results of Exercise 4.44, determine the number of periods following the step change for the expected value of the exponential smoothing statistic to be within 0.10 δ of the new time series level $\mu + \delta$. Plot the number of periods as a function of the smoothing constant. What conclusions can you draw?

4.59 Suppose that simple exponential smoothing is being used to forecast the process $y_t = \mu + \varepsilon_t$. At the start of period t^*, the mean of the process experiences a transient; that is, it shifts to a new level $\mu + \delta$, but reverts to its original level y at the start of the next period $t^* + 1$. The mean remains at this level for subsequent time periods. Show that the expected value of the exponentially smoothed statistic is

$$E(\hat{y}_t) = \begin{cases} \mu, & T \leq t^* \\ \mu + \delta\lambda(1-\lambda)^{T-t^*}, & T \geq t^* \end{cases}.$$

4.60 Using the results of Exercise 4.46, determine the number of periods that it will take following the impulse for the expected value of the exponential smoothing statistic to return to within 0.10 δ of the original time series level μ. Plot the number of periods as a function of the smoothing constant. What conclusions can you draw?

CHAPTER 5

AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

All models are wrong, some are useful.

GEORGE E. P. BOX, *British statistician*

5.1 INTRODUCTION

In the previous chapter, we discussed forecasting techniques that, in general, were based on some variant of exponential smoothing. The general assumption for these models was that any time series data can be represented as the sum of two distinct components: deterministic and stochastic (random). The former is modeled as a function of time whereas for the latter we assumed that some random noise that is added on to the deterministic signal generates the stochastic behavior of the time series. One very important assumption is that the random noise is generated through independent shocks to the process. In practice, however, this assumption is often violated. That is, usually successive observations show serial dependence. Under these circumstances, forecasting methods based on exponential smoothing may be inefficient and sometimes inappropriate

Introduction to Time Series Analysis and Forecasting, Second Edition.
Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci.

because they do not take advantage of the serial dependence in the observations in the most effective way. To formally incorporate this dependent structure, in this chapter we will explore a general class of models called autoregressive integrated moving average (MA) models or ARIMA models (also known as Box–Jenkins models).

5.2 LINEAR MODELS FOR STATIONARY TIME SERIES

In statistical modeling, we are often engaged in an endless pursuit of finding the ever elusive true relationship between certain inputs and the output. As cleverly put by the quote of this chapter, these efforts usually result in models that are nothing but approximations of the "true" relationship. This is generally due to the choices the analyst makes along the way to ease the modeling efforts. A major assumption that often provides relief in modeling efforts is the linearity assumption. A **linear filter**, for example, is a linear operation from one time series x_t to another time series y_t,

$$y_t = L(x_t) = \sum_{i=-\infty}^{+\infty} \psi_i x_{t-i} \tag{5.1}$$

with $t = \ldots, -1, 0, 1, \ldots$. In that regard the linear filter can be seen as a "process" that converts the input, x_t, into an output, y_t, and that conversion is not instantaneous but involves all (present, past, and future) values of the input in the form of a summation with different "weights", $\{\psi_i\}$, on each x_t. Furthermore, the linear filter in Eq. (5.1) is said to have the following properties:

1. **Time-invariant** as the coefficients $\{\psi_i\}$ do not depend on time.
2. **Physically realizable** if $\psi_i = 0$ for $i < 0$; that is, the output y_t is a linear function of the current and past values of the input: $y_t = \psi_0 x_t + \psi_1 x_{t-1} + \cdots$.
3. **Stable** if $\sum_{i=-\infty}^{+\infty} |\psi_i| < \infty$.

In linear filters, under certain conditions, some properties such as **stationarity** of the input time series are also reflected in the output. We discussed stationarity previously in Chapter 2. We will now give a more formal description of it before proceeding further with linear models for time series.

5.2.1 Stationarity

The **stationarity** of a time series is related to its statistical properties in time. That is, in the more strict sense, a stationary time series exhibits similar "statistical behavior" in time and this is often characterized as a constant probability distribution in time. However, it is usually satisfactory to consider the first two moments of the time series and define stationarity (or **weak stationarity**) as follows: (1) the expected value of the time series does not depend on time and (2) the autocovariance function defined as $\text{Cov}(y_t, y_{t+k})$ for any lag k is only a function of k and not time; that is, $\gamma_y(k) = \text{Cov}(y_t, y_{t+k})$.

In a crude way, the stationarity of a time series can be determined by taking arbitrary "snapshots" of the process at different points in time and observing the general behavior of the time series. If it exhibits "similar" behavior, one can then proceed with the modeling efforts under the assumption of stationarity. Further preliminary tests also involve observing the behavior of the autocorrelation function. A strong and slowly dying ACF will also suggest deviations from stationarity. Better and more methodological tests of stationarity also exist and we will discuss some of them later in this chapter. Figure 5.1 shows examples of stationary and nonstationary time series data.

5.2.2 Stationary Time Series

For a time-invariant and stable linear filter and a stationary input time series x_t with $\mu_x = E(x_t)$ and $\gamma_x(k) = \text{Cov}(x_t, x_{t+k})$, the output time series y_t given in Eq. (5.1) is also a stationary time series with

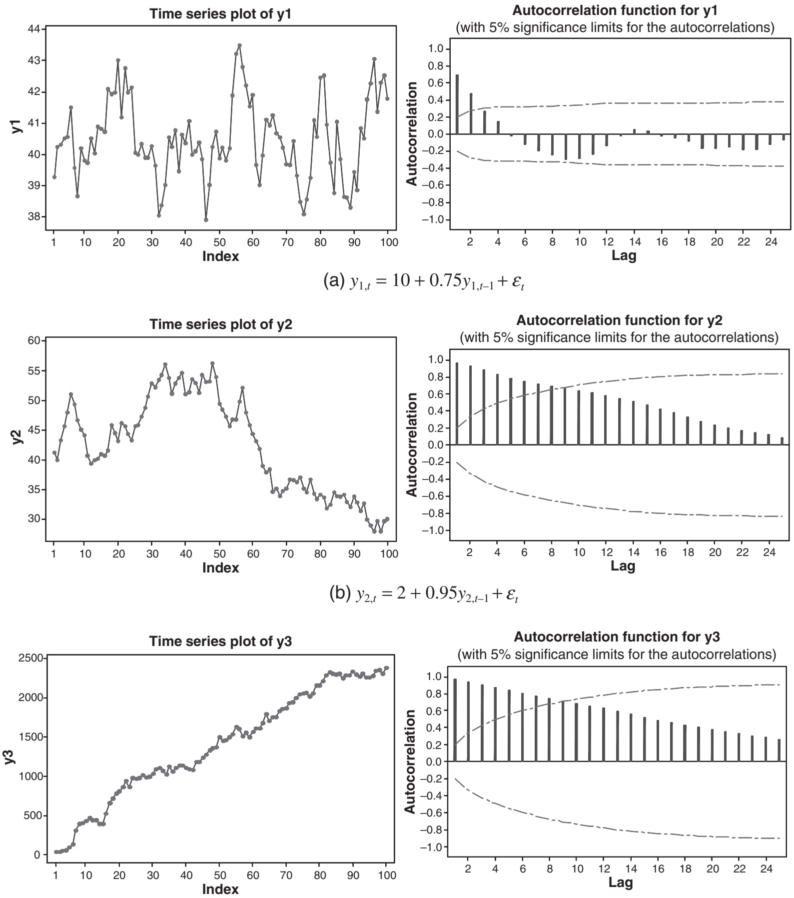
$$E(y_t) = \mu_y = \sum_{-\infty}^{\infty} \psi_i \mu_x$$

and

$$\text{Cov}(y_t, y_{t+k}) = \gamma_y(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_i \psi_j \gamma_x(i - j + k)$$

It is then easy to show that the following stable linear process with white noise time series, ε_t, is also stationary:

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{5.2}$$



(a) $y_{1,t} = 10 + 0.75y_{1,t-1} + \varepsilon_t$

(b) $y_{2,t} = 2 + 0.95y_{2,t-1} + \varepsilon_t$

(c) $y_{3,t} = 20 + y_{3,t-1} + \varepsilon_t$

FIGURE 5.1 Realizations of (a) stationary, (b) near nonstationary, and (c) nonstationary processes.

with $E(\varepsilon_t) = 0$, and

$$\gamma_\varepsilon(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{if } h \neq 0 \end{cases}$$

So for the autocovariance function of y_t, we have

$$\begin{aligned} \gamma_y(k) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_\varepsilon(i - j + k) \\ &= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \end{aligned} \tag{5.3}$$

We can rewrite the linear process in Eq. (5.2) in terms of the **backshift operator**, B, as

$$
\begin{aligned}
y_t &= \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \cdots \\
&= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t \\
&= \mu + \underbrace{\left(\sum_{i=0}^{\infty} \psi_i B^i \right)}_{=\Psi(\mathrm{B})} \varepsilon_t \\
&= \mu + \Psi(B)\, \varepsilon_t
\end{aligned}
\tag{5.4}
$$

This is called the **infinite moving average** and serves as a general class of models for any stationary time series. This is due to a theorem by Wold (1938) and basically states that **any** nondeterministic weakly stationary time series y_t can be represented as in Eq. (5.2), where $\{\psi_i\}$ satisfy $\sum_{i=0}^{\infty} \psi_i^2 < \infty$. A more intuitive interpretation of this theorem is that a stationary time series can be seen as the weighted sum of the present and past random "disturbances." For further explanations see Yule (1927) and Bisgaard and Kulahci (2005, 2011).

The theorem by Wold requires that the random shocks in (5.4) to be white noise which we defined as uncorrelated random shocks with constant variance. Some textbooks discuss independent or strong white noise for random shocks. It should be noted that there is a difference between correlation and independence. Independent random variables are also uncorrelated but the opposite is not always true. Independence between two random variables refers their joint probability distribution function being equal to the product of the marginal distributions. That is, two random variables X and Y are said to be independent if

$$
f(X, Y) = f_X(X) f_Y(Y)
$$

This can also be loosely interpreted as if X and Y are independent, knowing the value of X for example does not provide any information about what the value of Y might be.

For two uncorrelated random variables X and Y, we have their correlation and their covariance equal to zero. That is,

$$
\begin{aligned}
Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
&= E[XY] - E[X]E[Y] \\
&= 0
\end{aligned}
$$

This implies that if X and Y are uncorrelated, $E[XY] = E[X]E[Y]$.

Clearly if two random variables are independent, they are also uncorrelated since under independence we always have

$$\begin{aligned} E[XY] &= \iint xyf(x,y)dxdy \\ &= \iint xyf(x)f(y)dxdy \\ &= \left\{ \int xf(x)dx \right\} \left\{ \int yf(y)dx \right\} \\ &= E[X]E[Y] \end{aligned}$$

As we mentioned earlier, the opposite is not always true. To illustrate this with an example, consider X, a random variable with a symmetric probability density function around 0, i.e., $E[X] = 0$. Assume that the second variable Y is equal to $|X|$. Since knowing the value of X also determines the value of Y, these two variables are clearly not independent. However we can show that $E[Y] = 2\int_0^\infty xf(x)dx$ and $E[XY] = \int_0^\infty x^2 f(x)dx - \int_{-\infty}^0 x^2 f(x)dx = 0$ and hence $E[XY] = E[X]E[Y]$ This shows that X and Y are uncorrected but not independent.

Wold's decomposition theorem practically forms the foundation of the models we discuss in this chapter. This means that the strong assumption of independence is not necessarily needed except for the discussion on forecasting using ARIMA models in Section 5.8 where we assume the random shocks to be independent.

It can also be seen from Eq. (5.3) that there is a direct relation between the weights $\{\psi_i\}$ and the autocovariance function. In modeling a stationary time series as in Eq. (5.4), it is obviously impractical to attempt to estimate the infinitely many weights given in $\{\psi_i\}$. Although very powerful in providing a general representation of any stationary time series, the infinite moving average model given in Eq. (5.2) is useless in practice except for certain special cases:

1. Finite order moving average (MA) models where, except for a finite number of the weights in $\{\psi_i\}$, they are set to 0.
2. Finite order autoregressive (AR) models, where the weights in $\{\psi_i\}$ are generated using only a finite number of parameters.
3. A mixture of finite order autoregressive and moving average models (ARMA).

We shall now discuss each of these classes of models in great detail.

5.3 FINITE ORDER MOVING AVERAGE PROCESSES

In finite order moving average or MA models, conventionally ψ_0 is set to 1 and the weights that are not set to 0 are represented by the Greek letter θ with a minus sign in front. Hence a moving average process of order q (MA(q)) is given as

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \tag{5.5}$$

where $\{\varepsilon_t\}$ is white noise. Since Eq. (5.5) is a special case of Eq. (5.4) with only finite weights, an MA(q) process is **always** stationary regardless of values of the weights. In terms of the backward shift operator, the MA(q) process is

$$\begin{aligned} y_t &= \mu + \left(1 - \theta_1 B - \cdots - \theta_q B^q\right) \varepsilon_t \\ &= \mu + \left(1 - \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t \\ &= \mu + \Theta(B)\, \varepsilon_t \end{aligned} \tag{5.6}$$

where $\Theta(B) = 1 - \sum_{i=1}^{q} \theta_i B^i$.

Furthermore, since $\{\varepsilon_t\}$ is white noise, the expected value of the MA(q) process is simply

$$\begin{aligned} E(y_t) &= E\left(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}\right) \\ &= \mu \end{aligned} \tag{5.7}$$

and its variance is

$$\begin{aligned} \mathrm{Var}(y_t) = \gamma_y(0) &= \mathrm{Var}\left(\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}\right) \\ &= \sigma^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right) \end{aligned} \tag{5.8}$$

Similarly, the autocovariance at lag k can be calculated from

$$\begin{aligned} \gamma_y(k) &= \mathrm{Cov}(y_t, y_{t+k}) \\ &= E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q})(\varepsilon_{t+k} - \theta_1 \varepsilon_{t+k-1} - \cdots - \theta_q \varepsilon_{t+k-q})] \\ &= \begin{cases} \sigma^2(-\theta_k + \theta_1 \theta_{k+1} + \cdots + \theta_{q-k} \theta_q), & k = 1, 2, \ldots, q \\ 0, & k > q \end{cases} \end{aligned} \tag{5.9}$$

From Eqs. (5.8) and (5.9), the autocorrelation function of the MA(q) process is

$$\rho_y(k) = \frac{\gamma_y(k)}{\gamma_y(0)} = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2}, & k = 1, 2, \ldots, q \\ 0, \quad k > q \end{cases} \tag{5.10}$$

This feature of the ACF is very helpful in identifying the MA model and its appropriate order as it "cuts off" after lag q. In real life applications, however, the sample ACF, $r(k)$, will not necessarily be equal to zero after lag q. It is expected to become very small in absolute value after lag q. For a data set of N observations, this is often tested against $\pm 2/\sqrt{N}$ limits, where $1/\sqrt{N}$ is the approximate value for the standard deviation of the ACF for any lag under the assumption $\rho(k) = 0$ for all k's as discussed in Chapter 2.

Note that a more accurate formula for the standard error of the kth sample autocorrelation coefficient is provided by Bartlett (1946) as

$$s.e.\ (r(k)) = N^{-1/2}\left(1 + 2\sum_{j=1}^{k-1} r(j)^{*2}\right)^{1/2}$$

where

$$r(j)^* = \begin{cases} r(j) & \text{for } \rho(j) \neq 0 \\ 0 & \text{for } \rho(j) = 0 \end{cases}$$

A special case would be white noise data for which $\rho(j) = 0$ for all j's. Hence for a white noise process (i.e., no autocorrelation), a reasonable interval for the sample autocorrelation coefficients to fall in would be $\pm 2/\sqrt{N}$ and any indication otherwise may be considered as evidence for serial dependence in the process.

5.3.1 The First-Order Moving Average Process, MA(1)

The simplest finite order MA model is obtained when $q = 1$ in Eq. (5.5):

$$y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} \tag{5.11}$$

For the first-order moving average or MA(1) model, we have the autocovariance function as

$$\begin{aligned}\gamma_y(0) &= \sigma^2\left(1+\theta_1^2\right)\\ \gamma_y(1) &= -\theta_1\sigma^2\\ \gamma_y(k) &= 0, \quad k>1\end{aligned} \tag{5.12}$$

Similarly, we have the autocorrelation function as

$$\begin{aligned}\rho_y(1) &= \frac{-\theta_1}{1+\theta_1^2}\\ \rho_y(k) &= 0, \quad k>1\end{aligned} \tag{5.13}$$

From Eq. (5.13), we can see that the first lag autocorrelation in MA(1) is bounded as

$$\left|\rho_y(1)\right| = \frac{|\theta_1|}{1+\theta_1^2} \le \frac{1}{2} \tag{5.14}$$

and the autocorrelation function cuts off after lag 1.

Consider, for example, the following MA(1) model:

$$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

A realization of this model with its sample ACF is given in Figure 5.2. A visual inspection reveals that the mean and variance remain stable while there are some short runs where successive observations tend to follow each other for very brief durations, suggesting that there is indeed some positive autocorrelation in the data as revealed in the sample ACF plot.

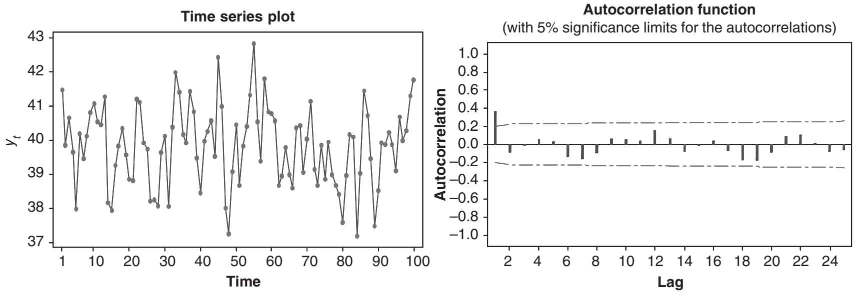


FIGURE 5.2 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$.

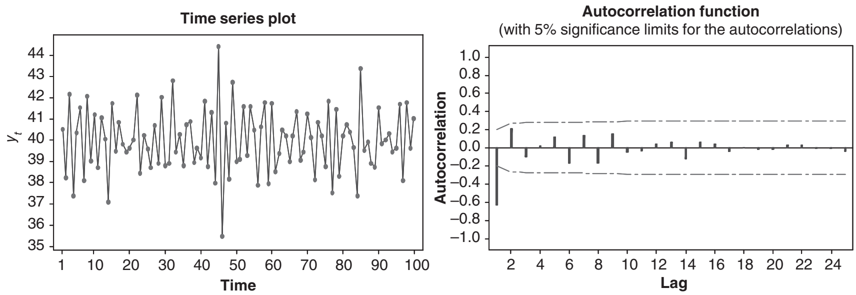


FIGURE 5.3 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$.

We can also consider the following model:

$$y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$$

A realization of this model is given in Figure 5.3. We can see that observations tend to oscillate successively. This suggests a negative autocorrelation as confirmed by the sample ACF plot.

5.3.2 The Second-Order Moving Average Process, MA(2)

Another useful finite order moving average process is MA(2), given as

$$\begin{aligned} y_t &= \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} \\ &= \mu + \left(1 - \theta_1 B - \theta_2 B^2\right)\varepsilon_t \end{aligned} \tag{5.15}$$

The autocovariance and autocorrelation functions for the MA(2) model are given as

$$\begin{aligned} \gamma_y(0) &= \sigma^2\left(1 + \theta_1^2 + \theta_2^2\right) \\ \gamma_y(1) &= \sigma^2(-\theta_1 + \theta_1\theta_2) \\ \gamma_y(2) &= \sigma^2(-\theta_2) \\ \gamma_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \rho_y(1) &= \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(2) &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \rho_y(k) &= 0, \quad k > 2 \end{aligned} \tag{5.17}$$

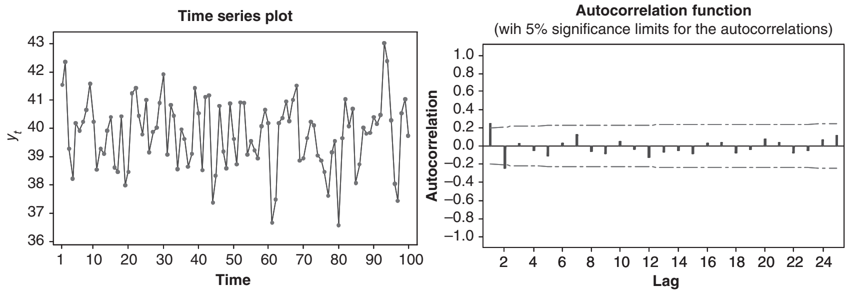


FIGURE 5.4 A realization of the MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$.

Figure 5.4 shows the time series plot and the autocorrelation function for a realization of the MA(2) model:

$$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$$

Note that the sample ACF cuts off after lag 2.

5.4 FINITE ORDER AUTOREGRESSIVE PROCESSES

As mentioned in Section 5.1, while it is quite powerful and important, Wold's decomposition theorem does not help us much in our modeling and forecasting efforts as it implicitly requires the estimation of the infinitely many weights, $\{\psi_i\}$. In Section 5.2 we discussed a special case of this decomposition of the time series by assuming that it can be adequately modeled by only estimating a finite number of weights and setting the rest equal to 0. Another interpretation of the finite order MA processes is that at any given time, of the infinitely many past disturbances, only a finite number of those disturbances "contribute" to the current value of the time series and that the time window of the contributors "moves" in time, making the "oldest" disturbance obsolete for the next observation. It is indeed not too far fetched to think that some processes might have these intrinsic dynamics. However, for some others, we may be required to consider the "lingering" contributions of the disturbances that happened back in the past. This will of course bring us back to square one in terms of our efforts in estimating infinitely many weights. Another solution to this problem is through the autoregressive models in which the infinitely many weights are assumed to follow a distinct pattern and can be successfully represented with only a handful of parameters. We shall now consider some special cases of autoregressive processes.

5.4.1 First-Order Autoregressive Process, AR(1)

Let us first consider again the time series given in Eq. (5.2):

$$\begin{aligned} y_t &= \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t \\ &= \mu + \Psi(B)\,\varepsilon_t \end{aligned}$$

where $\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$. As in the finite order MA processes, one approach to modeling this time series is to assume that the contributions of the disturbances that are way in the past should be small compared to the more recent disturbances that the process has experienced. Since the disturbances are independently and identically distributed random variables, we can simply assume a set of infinitely many weights in descending magnitudes reflecting the diminishing magnitudes of contributions of the disturbances in the past. A simple, yet intuitive set of such weights can be created following an exponential decay pattern. For that we will set $\psi_i = \phi^i$, where $|\phi| < 1$ to guarantee the exponential "decay." In this notation, the weights on the disturbances starting from the current disturbance and going back in past will be $1, \phi, \phi^2, \phi^3, \ldots$ Hence Eq. (5.2) can be written as

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots \\ &= \mu + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \end{aligned} \tag{5.18}$$

From Eq. (5.18), we also have

$$y_{t-1} = \mu + \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \cdots \tag{5.19}$$

We can then combine Eqs. (5.18) and (5.19) as

$$\begin{aligned} y_t &= \mu + \varepsilon_t + \underbrace{\phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots}_{=\phi y_{t-1} - \phi\mu} \\ &= \underbrace{\mu - \phi\mu}_{=\delta} + \phi y_{t-1} + \varepsilon_t \\ &= \delta + \phi y_{t-1} + \varepsilon_t \end{aligned} \tag{5.20}$$

where $\delta = (1 - \phi)\,\mu$. The process in Eq. (5.20) is called a **first-order autoregressive process**, AR(1), because Eq. (5.20) can be seen as a regression of y_t on y_{t-1} and hence the term **auto**regressive process.

The assumption of $|\phi| < 1$ results in the weights that decay exponentially in time and also guarantees that $\sum_{i=0}^{+\infty} |\psi_i| < \infty$. This means that an AR(1) process is stationary if $|\phi| < 1$. For $|\phi| > 1$, past disturbances will get exponentially increasing weights as time goes on and the resulting time series will be explosive. Box et al. (2008) argue that this type of processes are of little practical interest and therefore only consider cases where $|\phi| = 1$ and $|\phi| < 1$. The solution in (5.18) does indeed not converge for $|\phi| > 1$. We can however rewrite the AR(1) process for y_{t+1}

$$y_{t+1} = \phi y_t + a_{t+1} \tag{5.21}$$

For y_t, we then have

$$\begin{aligned} y_t &= -\phi^{-1}\mu + \phi^{-1} y_{t+1} - \phi^{-1}\varepsilon_{t+1} \\ &= -\phi^{-1}\mu + \phi^{-1}\left(-\phi^{-1}\mu + \phi^{-1} y_{t+2} - \phi^{-1}\varepsilon_{t+2}\right) - \phi^{-1}\varepsilon_{t+1} \\ &= -\left(\phi^{-1} + \phi^{-2}\right)\mu + \phi^{-2} y_{t+2} - \phi^{-1}\varepsilon_{t+1} - \phi^{-2}\varepsilon_{t+2} \\ &\ \vdots \\ &= -\mu \sum_{i=1}^{\infty} \phi^{-1} - \sum_{i=1}^{\infty} \phi^{-1}\varepsilon_{t+1} \end{aligned} \tag{5.22}$$

For $|\phi| > 1$ we have $|\phi^{-1}| < 1$ and therefore the solution for y_t given in (5.22) is stationary. The only problem is that it involves future values of disturbances. This of course is impractical as this type of models requires knowledge about the future to make forecasts about it. These are called non-causal models. Therefore there exists a stationary solution for an AR(1) process when $|\phi| > 1$, however, it results in a non-causal model. Throughout the book when we discuss the stationary autoregressive models, we implicitly refer to the causal autoregressive models. We can in fact show that an AR(I) process is nonstationary if and only if $|\phi| = 1$.

The mean of a stationary AR(1) process is

$$E(y_t) = \mu = \frac{\delta}{1 - \phi} \tag{5.23}$$

The autocovariance function of a stationary AR(1) can be calculated from Eq. (5.18) as

$$\gamma(k) = \sigma^2 \phi^k \frac{1}{1-\phi^2} \quad \text{for } k = 0, 1, 2, \ldots \tag{5.24}$$

The covariance is then given as

$$\gamma(0) = \sigma^2 \frac{1}{1-\phi^2} \tag{5.25}$$

Correspondingly, the autocorrelation function for a stationary AR(1) process is given as

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k \quad \text{for } k = 0, 1, 2, \ldots \tag{5.26}$$

Hence the ACF for an AR(1) process has an exponential decay form.

A realization of the following AR(1) model,

$$y_t = 8 + 0.8 y_{t-1} + \varepsilon_t$$

is shown in Figure 5.5. As in the MA(1) model with $\theta = -0.8$, we can observe some short runs during which observations tend to move in the upward or downward direction. As opposed to the MA(1) model, however, the duration of these runs tends to be longer and the trend tends to linger. This can also be observed in the sample ACF plot.

Figure 5.6 shows a realization of the AR(1) model $y_t = 8 - 0.8 y_{t-1} + \varepsilon_t$. We observe that instead of lingering runs, the observations exhibit jittery up/down movements because of the negative ϕ value.

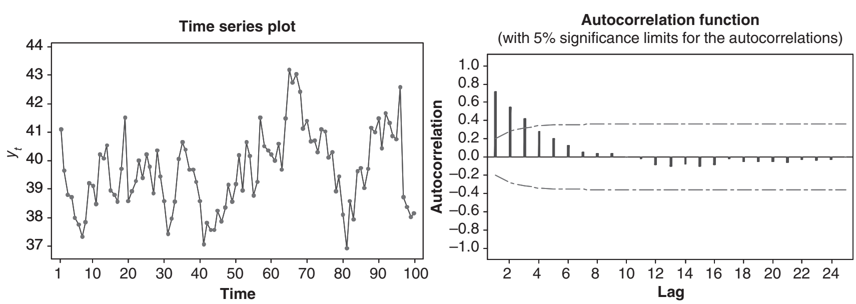


FIGURE 5.5 A realization of the AR(1) process, $y_t = 8 + 0.8 y_{t-1} + \varepsilon_t$.

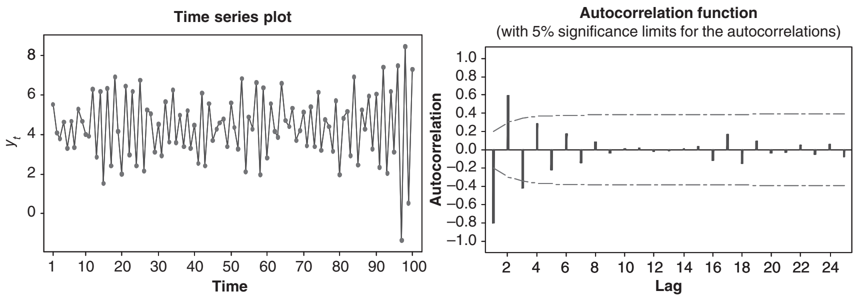


FIGURE 5.6 A realization of the AR(1) process, $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$.

5.4.2 Second-Order Autoregressive Process, AR(2)

In this section, we will first start with the obvious extension of Eq. (5.20) to include the observation y_{t-2} as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \tag{5.27}$$

We will then show that Eq. (5.27) can be represented in the infinite MA form and provide the conditions of stationarity for y_t in terms of ϕ_1 and ϕ_2. For that we will rewrite Eq. (5.27) as

$$(1 - \phi_1 B - \phi_2 B^2)y_t = \delta + \varepsilon_t \tag{5.28}$$

or

$$\Phi(B)y_t = \delta + \varepsilon_t \tag{5.29}$$

Furthermore, applying $\Phi(B)^{-1}$ to both sides, we obtain

$$\begin{aligned} y_t &= \underbrace{\Phi(B)^{-1}\,\delta}_{=\mu} + \underbrace{\Phi(B)^{-1}}_{=\Psi(B)}\,\varepsilon_t \\ &= \mu + \Psi(B)\,\varepsilon_t \\ &= \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \psi_i B^i \varepsilon_t \end{aligned} \tag{5.30}$$

where

$$\mu = \Phi(B)^{-1}\,\delta \tag{5.31}$$

and

$$\Phi(B)^{-1} = \sum_{i=0}^{\infty} \psi_i B^i = \Psi(B) \tag{5.32}$$

We can use Eq. (5.32) to obtain the weights in Eq. (5.30) in terms of ϕ_1 and ϕ_2. For that, we will use

$$\Phi(B)\,\Psi(B) = 1 \tag{5.33}$$

That is,

$$(1 - \phi_1 B - \phi_2 B^2)(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots) = 1$$

or

$$\begin{aligned} &\psi_0 + (\psi_1 - \phi_1\psi_0)B + (\psi_2 - \phi_1\psi_1 - \phi_2\psi_0)B^2 \\ &\quad + \cdots + (\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2})B^j + \cdots = 1 \end{aligned} \tag{5.34}$$

Since on the right-hand side of the Eq. (5.34) there are no backshift operators, for $\Phi(B)\,\Psi(B) = 1$, we need

$$\begin{aligned} \psi_0 &= 1 \\ (\psi_1 - \phi_1\psi_0) &= 0 \\ (\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2}) &= 0 \quad \text{for all } j = 2, 3, \ldots \end{aligned} \tag{5.35}$$

The equations in (5.35) can indeed be solved for each ψ_j in a futile attempt to estimate infinitely many parameters. However, it should be noted that the ψ_j in Eq. (5.35) satisfy the second-order linear difference equation and that they can be expressed as the solution to this equation in terms of the two roots m_1 and m_2 of the associated polynomial

$$m^2 - \phi_1 m - \phi_2 = 0 \tag{5.36}$$

If the roots obtained by

$$m_1, m_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

satisfy $|m_1|, |m_2| < 1$, then we have $\sum_{i=0}^{+\infty} |\psi_i| < \infty$. Hence if the roots m_1 and m_2 are both less than 1 in absolute value, then the AR(2) model is causal and stationary. Note that if the roots of Eq. (5.36) are complex conjugates of the form $a \pm ib$, the condition for stationarity is that $\sqrt{a^2 + b^2} < 1$. Furthermore, under the condition that $|m_1|, |m_2| < 1$, the AR(2) time series, $\{y_t\}$, has an infinite MA representation as in Eq. (5.30).

This implies that for the second-order autoregressive process to be stationary, the parameters ϕ_1 and ϕ_2 must satisfy.

$$\begin{aligned} &\phi_1 + \phi_2 < 1 \\ &\phi_2 - \phi_1 < 1 \\ &|\phi_2| < 1 \end{aligned}$$

Now that we have established the conditions for the stationarity of an AR(2) time series, let us now consider its mean, autocovariance, and autocorrelation functions. From Eq. (5.27), we have

$$\begin{aligned} E(y_t) &= \delta + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + 0 \\ \mu &= \delta + \phi_1 \mu + \phi_2 \mu \\ \Rightarrow \mu &= \frac{\delta}{1 - \phi_1 - \phi_2} \end{aligned} \tag{5.37}$$

Note that for $1 - \phi_1 - \phi_2 = 0$, $m = 1$ is one of the roots for the associated polynomial in Eq. (5.36) and hence the time series is deemed nonstationary. The autocovariance function is

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(y_t, y_{t-k}) \\ &= \mathrm{Cov}(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, y_{t-k}) \\ &= \phi_1 \mathrm{Cov}(y_{t-1}, y_{t-k}) + \phi_2 \mathrm{Cov}(y_{t-2}, y_{t-k}) + \mathrm{Cov}(\varepsilon_t, y_{t-k}) \\ &= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k > 0 \end{cases} \end{aligned} \tag{5.38}$$

Thus $\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2$ and

$$\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2), \quad k = 1, 2, \ldots \tag{5.39}$$

The equations in (5.39) are called the **Yule–Walker** equations for $\gamma(k)$. Similarly, we can obtain the autocorrelation function by dividing Eq. (5.39) by $\gamma(0)$:

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2), \quad k = 1, 2, \ldots \tag{5.40}$$

The Yule–Walker equations for $\rho(k)$ in Eq. (5.40) can be solved recursively as

$$\begin{aligned}
\rho(1) &= \phi_1 \underbrace{\rho(0)}_{=1} + \phi_2 \underbrace{\rho(-1)}_{=\rho(1)} \\
&= \frac{\phi_1}{1-\phi_2} \\
\rho(2) &= \phi_1\rho(1) + \phi_2 \\
\rho(3) &= \phi_1\rho(2) + \phi_2\rho(1) \\
&\vdots
\end{aligned}$$

A general solution can be obtained through the roots m_1 and m_2 of the associated polynomial $m^2 - \phi_1 m - \phi_2 = 0$. There are three cases.

Case 1. If m_1 and m_2 are distinct, real roots, we then have

$$\rho(k) = c_1 m_1^k + c_2 m_2^k, \quad k = 0, 1, 2, \ldots \tag{5.41}$$

where c_1 and c_2 are particular constants and can, for example, be obtained from $\rho(0)$ and $\rho(1)$. Moreover, since for stationarity we have $\left|m_1\right|, \left|m_2\right| < 1$, in this case, the autocorrelation function is a **mixture of two exponential decay terms**.

Case 2. If m_1 and m_2 are complex conjugates in the form of $a \pm ib$, we then have

$$\rho(k) = R^k\left[c_1\cos(\lambda k) + c_2\sin(\lambda k)\right], \quad k = 0, 1, 2, \ldots \tag{5.42}$$

where $R = |m_i| = \sqrt{a^2 + b^2}$ and λ is determined by $\cos(\lambda) = a/R$, $\sin(\lambda) = b/R$. Hence we have $a \pm ib = R[\cos(\lambda) \pm i\sin(\lambda)]$. Once again c_1 and c_2 are particular constants. The ACF in this case has the form of a **damped sinusoid**, with damping factor R and frequency λ; that is, the period is $2\pi/\lambda$.

Case 3. If there is one real root m_0, $m_1 = m_2 = m_0$, we then have

$$\rho(k) = (c_1 + c_2 k) m_0^k \quad k = 0, 1, 2, \ldots \tag{5.43}$$

In this case, the ACF will exhibit an exponential decay pattern.

In case 1, for example, an AR(2) model can be seen as an "adjusted" AR(1) model for which a single exponential decay expression as in the AR(1) model is not enough to describe the pattern in the ACF, and hence an additional exponential decay expression is "added" by introducing the second lag term, y_{t-2}.

Figure 5.7 shows a realization of the AR(2) process

$$y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$$

Note that the roots of the associated polynomial of this model are real. Hence the ACF is a mixture of two exponential decay terms.

Similarly, Figure 5.8 shows a realization of the following AR(2) process

$$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t.$$

For this process, the roots of the associated polynomial are complex conjugates. Therefore the ACF plot exhibits a damped sinusoid behavior.

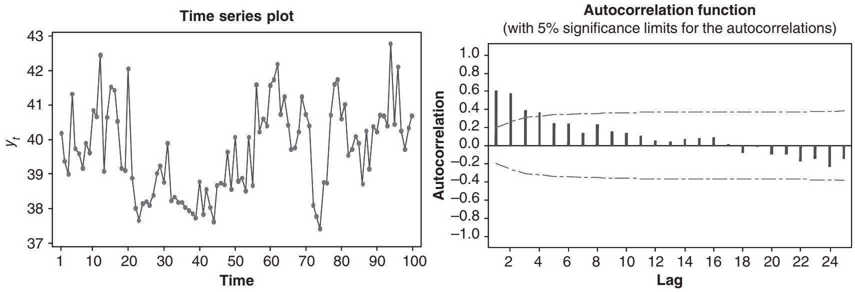


FIGURE 5.7 A realization of the AR(2) process, $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$.

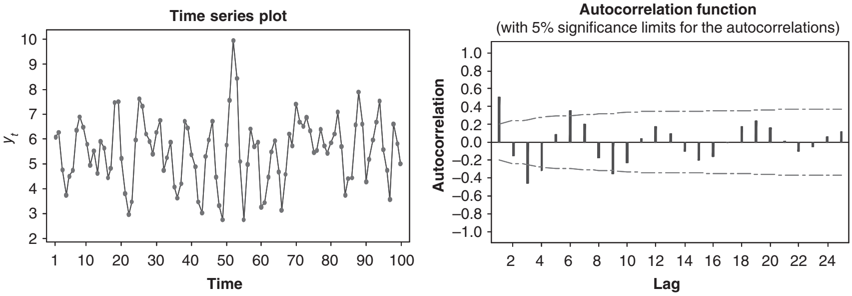


FIGURE 5.8 A realization of the AR(2) process, $y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$.

5.4.3 General Autoregressive Process, AR(p)

From the previous two sections, a general, pth-order AR model is given as

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t \tag{5.44}$$

where ε_t is white noise. Another representation of Eq. (5.44) can be given as

$$\Phi(B) y_t = \delta + \varepsilon_t \tag{5.45}$$

where $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$.

The AR(p) time series $\{y_t\}$ in Eq. (5.44) is causal and stationary if the roots of the associated polynomial

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \cdots - \phi_p = 0 \tag{5.46}$$

are less than one in absolute value. Furthermore, under this condition, the AR(p) time series $\{y_t\}$ is also said to have an **absolutely summable** infinite MA representation

$$y_t = \mu + \Psi(B)\,\varepsilon_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \tag{5.47}$$

where $\Psi(B) = \Phi(B)^{-1}$ with $\sum_{i=0}^{\infty} |\psi_i| < \infty$.

As in AR(2), the weights of the random shocks in Eq. (5.47) can be obtained from $\Phi(B)\Psi(B) = 1$ as

$$\begin{aligned} \psi_j &= 0, \quad j < 0 \\ \psi_0 &= 1 \\ \psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} - \cdots - \phi_p\psi_{j-p} &= 0 \quad \text{for all } j = 1, 2, \ldots \end{aligned} \tag{5.48}$$

We can easily show that, for the stationary AR(p) process

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

and

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(y_t, y_{t-k}) \\ &= \mathrm{Cov}(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t, y_{t-k}) \\ &= \sum_{i=1}^{p} \phi_i \mathrm{Cov}(y_{t-i}, y_{t-k}) + \mathrm{Cov}(\varepsilon_t, y_{t-k}) \\ &= \sum_{i=1}^{p} \phi_i \gamma(k-i) + \begin{cases} \sigma^2 & \text{if } k = 0 \\ 0 & \text{if } k > 0 \end{cases} \end{aligned} \tag{5.49}$$

Thus we have

$$\gamma(0) = \sum_{i=1}^{p} \phi_i \gamma(i) + \sigma^2 \tag{5.50}$$

$$\Rightarrow \gamma(0)\left[1 - \sum_{i=1}^{p} \phi_i \rho(i)\right] = \sigma^2 \tag{5.51}$$

By dividing Eq. (5.49) by $\gamma(0)$ for $k > 0$, it can be observed that the ACF of an AR(p) process satisfies the Yule–Walker equations

$$\rho(k) = \sum_{i=1}^{p} \phi_i \rho(k-i), \quad k = 1, 2, \ldots \tag{5.52}$$

The equations in (5.52) are pth-order **linear difference equations**, implying that the ACF for an AR(p) model can be found through the p roots of the associated polynomial in Eq. (5.46). For example, if the roots are all distinct and real, we have

$$\rho(k) = c_1 m_1^k + c_2 m_2^k + \cdots + c_p m_p^k, \quad k = 1, 2, \ldots \tag{5.53}$$

where $c_1, c_2, \ldots, c_p$ are particular constants. However, in general, the roots may not all be distinct or real. Thus the ACF of an AR(p) process can be a **mixture of exponential decay and damped sinusoid** expressions depending on the roots of Eq. (5.46).

5.4.4 Partial Autocorrelation Function, PACF

In Section 5.2, we saw that the ACF is an excellent tool in identifying the order of an MA(q) process, because it is expected to "cut off" after lag q. However, in the previous section, we pointed out that the ACF is not as useful in the identification of the order of an AR(p) process for which it will most likely have a mixture of exponential decay and damped sinusoid expressions. Hence such behavior, while indicating that the process might have an AR structure, fails to provide further information about the order of such structure. For that, we will define and employ the **partial autocorrelation function** (PACF) of the time series. But before that, we discuss the concept of partial correlation to make the interpretation of the PACF easier.

Partial Correlation Consider three random variables X, Y, and Z. Then consider simple linear regression of X on Z and Y on Z as

$$\hat{X} = a_1 + b_1 Z \quad \text{where } b_1 = \frac{\text{Cov}(Z, X)}{\text{Var}(Z)}$$

and

$$\hat{Y} = a_2 + b_2 Z \quad \text{where } b_2 = \frac{\text{Cov}(Z, Y)}{\text{Var}(Z)}$$

Then the errors can be obtained from

$$X^* = X - \hat{X} = X - (a_1 + b_1 Z)$$

and

$$Y^* = Y - \hat{Y} = Y - (a_2 + b_2 Z)$$

Then the **partial correlation** between X and Y after adjusting for Z is defined as the correlation between X^* and Y^*; $\text{corr}(X^*, Y^*) = \text{corr}(X - \hat{X}, Y - \hat{Y})$. That is, partial correlation can be seen as the correlation between two variables after being adjusted for a common factor that may be affecting them. The generalization is of course possible by allowing for adjustment for more than just one factor.

Partial Autocorrelation Function Following the above definition, the **PACF** between y_t and y_{t-k} is the autocorrelation between y_t and y_{t-k} after adjusting for $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$. Hence for an AR($p$) model the PACF between y_t and y_{t-k} for $k > p$ should be equal to zero. A more formal definition can be found below.

Consider a stationary time series model $\{y_t\}$ that is not necessarily an AR process. Further consider, for any fixed value of k, the Yule–Walker equations for the ACF of an AR(p) process given in Eq. (5.52) as

$$\rho(j) = \sum_{i=1}^{k} \phi_{ik} \rho(j-i), \quad j = 1, 2, \dots, k \tag{5.54}$$

or

$$\begin{aligned}
\rho(1) &= \phi_{1k} + \phi_{2k}\rho(1) + \cdots + \phi_{kk}\rho(k-1) \\
\rho(2) &= \phi_{1k}\rho(1) + \phi_{2k} + \cdots + \phi_{kk}\rho(k-2) \\
&\;\;\vdots \\
\rho(k) &= \phi_{1k}\rho(k-) + \phi_{2k}\rho(k-2) + \cdots + \phi_{kk}
\end{aligned}$$

Hence we can write the equations in (5.54) in matrix notation as

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix} \tag{5.55}$$

or

$$\mathbf{P}_k \phi_k = \rho_k \tag{5.56}$$

where

$$\mathbf{P}_k = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(3) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix},$$

$$\phi_k = \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \phi_{3k} \\ \vdots \\ \phi_{kk} \end{bmatrix}, \quad \text{and} \quad \rho_k = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(k) \end{bmatrix}.$$

Thus to solve for ϕ_k, we have

$$\phi_k = \mathbf{P}_k^{-1} \rho_k \tag{5.57}$$

For any given k, $k = 1, 2, \dots$, the last coefficient ϕ_{kk} is called the partial autocorrelation of the process at lag k. Note that for an AR(p) process $\phi_{kk} = 0$ for $k > p$. Hence we say that the PACF cuts off after lag p for an AR(p). This suggests that the PACF can be used in identifying the order of an AR process similar to how the ACF can be used for an MA process.

For sample calculations, $\hat{\phi}_{kk}$, the sample estimate of ϕ_{kk}, is obtained by using the sample ACF, $r(k)$. Furthermore, in a sample of N observations from an AR(p) process, $\hat{\phi}_{kk}$ for $k > p$ is approximately normally distributed with

$$E(\hat{\phi}_{kk}) \approx 0 \quad \text{and} \quad \text{Var}(\hat{\phi}_{kk}) \approx \frac{1}{N} \tag{5.58}$$

Hence the 95% limits to judge whether any $\hat{\phi}_{kk}$ is statistically significantly different from zero are given by $\pm 2/\sqrt{N}$. For further detail see Quenouille (1949), Jenkins (1954, 1956), and Daniels (1956).

Figure 5.9 shows the sample PACFs of the models we have considered so far. In Figure 5.9a we have the sample PACF of the realization of the MA(1) model with $\theta = 0.8$ given in Figure 5.3. It exhibits an exponential decay pattern. Figure 5.9b shows the sample PACF of the realization of the MA(2) model in Figure 5.4 and it also has an exponential decay pattern in absolute value since for this model the roots of the associated polynomial are real. Figures 5.9c and 5.9d show the sample PACFs of the realization of the AR(1) model with $\phi = 0.8$ and $\phi = -0.8$, respectively. In both

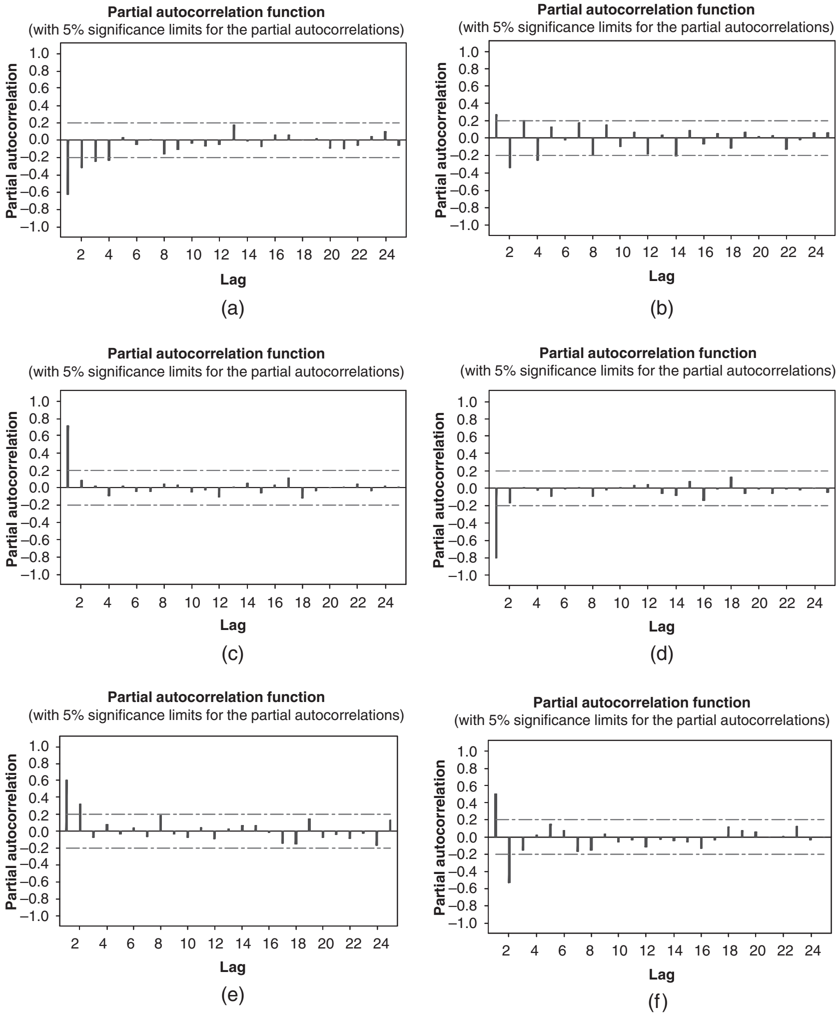


FIGURE 5.9 Partial autocorrelation functions for the realizations of (a) MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$; (b) MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$; (c) AR(1) process, $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$; (d) AR(1) process, $y_t = 8 - 0.8y_{t-1} + \varepsilon_t$; (e) AR(2) process, $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$; and (f) AR(2) process, $y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$.

cases the PACF “cuts off” after the first lag. That is, the only significant sample PACF value is at lag 1, suggesting that the AR(1) model is indeed appropriate to fit the data. Similarly, in Figures 5.9e and 5.9f, we have the sample PACFs of the realizations of the AR(2) model. Note that the sample PACF cuts off after lag 2.

As we discussed in Section 5.3, finite order MA processes are stationary. On the other hand as in the causality concept we discussed for the autoregressive processes, we will impose some restrictions on the parameters of the MA models as well. Consider for example the MA(1) model in (5.11)

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} \tag{5.59}$$

Note that for the sake of simplicity, in (5.59) we consider a centered process, i.e. $E(y_t) = 0$.

We can then rewrite (5.59) as

$$\begin{aligned}
\varepsilon_t &= y_t + \theta_1 \varepsilon_{t-1} \\
&= y_t + \theta_1 \left[y_{t-1} + \theta_1 \varepsilon_{t-2} \right] \\
&= y_t + \theta_1 y_{t-1} + \theta_1^2 \varepsilon_{t-2} \\
&\vdots \\
&= \sum_{i=0}^{\infty} \theta_1^i y_{t-i}
\end{aligned} \tag{5.60}$$

It can be seen from (5.60) that for $|\theta_1| < 1$, ε_t is a convergent series of current and past observations and the process is called an **invertible** moving average process. Similar to the causality argument, for $|\theta_1| > 1$, ε_t can be written as a convergent series of future observations and is called noninvertible. When $|\theta_1| = 1$, the MA(1) process is considered noninvertible in a more restricted sense (Brockwell and Davis (1991)).

The direct implication of invertibility becomes apparent in model identification. Consider the MA(1) process as an example. The first lag autocorrelation for that process is given as

$$\rho(1) = \frac{-\theta_1}{1 + \theta_1^2} \tag{5.61}$$

This allows for the calculation of θ_1 for a given $\rho(1)$ by rearranging (5.61) as

$$\theta_1^2 - \frac{\theta_1}{\rho(1)} + 1 = 0 \tag{5.62}$$

and solving for θ_1. Except for the case of a repeated root, this equation has two solutions. Consider for example $\rho(1) = 0.4$ for which both $\theta_1 = 0.5$ and $\theta_1 = 2$ are the solutions for (5.62). Following the above argument, only $\theta_1 = 0.5$ yields the invertible MA(1) process. It can be shown that when there are multiple solutions for possible values of MA parameters, there is only one solution that will satisfy the invertibility condition (Box et al. (2008), Section 6.4.1).

Consider the MA(q) process

$$\begin{aligned} y_t &= \mu + \left(1 - \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t \\ &= \mu + \Theta(B)\, \varepsilon_t \end{aligned}$$

After multiplying both sides with $\Theta(B)^{-1}$, we have

$$\begin{aligned} \Theta(B)^{-1}\, y_t &= \Theta(B)^{-1}\, \mu + \varepsilon_t \\ \Pi(B)\, y_t &= \delta + \varepsilon_t \end{aligned} \tag{5.63}$$

where $\Pi(B) = 1 - \sum_{i=1}^{\infty} \pi_i B^i = \Theta(B)^{-1}$ and $\Theta(B)^{-1}\, \mu = \delta$. Hence the infinite AR representation of an MA(q) process is given as

$$y_t - \sum_{i=1}^{\infty} \pi_i y_{t-i} = \delta + \varepsilon_t \tag{5.64}$$

with $\sum_{i=1}^{\infty} \left|\pi_i\right| < \infty$. The π_i can be determined from

$$(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)(1 - \pi_1 B - \pi_2 B^2 + \cdots) = 1 \tag{5.65}$$

which in turn yields

$$\begin{aligned} \pi_1 + \theta_1 &= 0 \\ \pi_2 - \theta_1 \pi_1 + \theta_2 &= 0 \\ &\vdots \\ \pi_j - \theta_1 \pi_{j-1} - \cdots - \theta_q \pi_{j-q} &= 0 \end{aligned} \tag{5.66}$$

with $\pi_0 = -1$ and $\pi_j = 0$ for $j < 0$. Hence as in the previous arguments for the stationarity of AR(p) models, the π_i are the solutions to the qth-order linear difference equations and therefore the condition for the invertibility of an MA(q) process turns out to be very similar to the stationarity condition of an AR(p) process: the roots of the associated polynomial given in Eq. (5.66) should be less than 1 in absolute value,

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - \cdots - \theta_q = 0 \tag{5.67}$$

An invertible MA(q) process can then be written as an infinite AR process.

Correspondingly, for such a process, adjusting for $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ does not necessarily eliminate the correlation between y_t and y_{t-k} and therefore its PACF will never "cut off." In general, the PACF of an MA(q) process is a **mixture of exponential decay and damped sinusoid** expressions.

The ACF and the PACF do have very distinct and indicative properties for MA and AR models, respectively. Therefore, in model identification, we strongly recommend the use of both the sample ACF and the sample PACF **simultaneously**.

5.5 MIXED AUTOREGRESSIVE–MOVING AVERAGE PROCESSES

In the previous sections we have considered special cases of Wold's decomposition of a stationary time series represented as a weighted sum of infinite random shocks. In an AR(1) process, for example, the weights in the infinite sum are forced to follow an exponential decay form with ϕ as the rate of decay. Since there are no restrictions apart from $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ on the weights (ψ_i), it may not be possible to approximate them by an exponential decay pattern. For that, we will need to increase the order of the AR model to approximate any pattern that these weights may in fact be exhibiting. On some occasions, however, it is possible to make simple adjustments to the exponential decay pattern by adding only a few terms and hence to have a more parsimonious model. Consider, for example, that the weights ψ_i do indeed exhibit an exponential decay pattern with a constant rate except for the fact that ψ_1 is not equal to this rate of decay as it would be in the case of an AR(1) process. Hence instead of increasing the order of the AR model to accommodate for this "anomaly," we can add an MA(1) term that will simply adjust ψ_1 while having no effect on the rate of exponential decay pattern of the rest of the weights. This results in a mixed **autoregressive moving average** or ARMA(1,1) model. In general, an ARMA(p, q) model is given as

$$\begin{aligned} y_t &= \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\ &\quad - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \\ &= \delta + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \end{aligned} \tag{5.68}$$

or

$$\Phi(B)\, y_t = \delta + \Theta(B)\, \varepsilon_t \tag{5.69}$$

where ε_t is a white noise process.

5.5.1 Stationarity of ARMA(*p, q*) Process

The **stationarity** of an ARMA process is related to the AR component in the model and can be checked through the roots of the associated polynomial

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - \cdots - \phi_p = 0. \tag{5.70}$$

If all the roots of Eq. (5.70) are less than one in absolute value, then ARMA(p, q) is stationary. This also implies that, under this condition, ARMA(p, q) has an infinite MA representation as

$$y_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \mu + \Psi(B)\, \varepsilon_t \tag{5.71}$$

with $\Psi(B) = \Phi(B)^{-1}\, \Theta(B)$. The coefficients in $\Psi(B)$ can be found from

$$\psi_i - \phi_1 \psi_{i-1} - \phi_2 \psi_{i-2} - \cdots - \phi_p \psi_{i-p} = \begin{cases} -\theta_i, & i = 1, \ldots, q \\ 0, & i > q \end{cases} \tag{5.72}$$

and $\psi_0 = 1$.

5.5.2 Invertibility of ARMA(*p, q*) Process

Similar to the stationarity condition, the **invertibility** of an ARMA process is related to the MA component and can be checked through the roots of the associated polynomial

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - \cdots - \theta_q = 0 \tag{5.73}$$

If all the roots of Eq. (5.71) are less than one in absolute value, then ARMA(p, q) is said to be invertible and has an infinite AR representation,

$$\Pi(B) y_t = \alpha + \varepsilon_t \tag{5.74}$$

where $\alpha = \Theta(B)^{-1}\,\delta$ and $\Pi(B) = \Theta(B)^{-1}\,\Phi(B)$. The coefficients in $\Pi(B)$ can be found from

$$\pi_i - \theta_1\pi_{i-1} - \theta_2\pi_{i-2} - \cdots - \theta_q\pi_{i-q} = \begin{cases} \phi_i, & i = 1, \ldots, p \\ 0, & i > p \end{cases} \tag{5.75}$$

and $\pi_0 = -1$.

In Figure 5.10 we provide realizations of two ARMA(1,1) models:

$$y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1} \quad \text{and} \quad y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}.$$

Note that the sample ACFs and PACFs exhibit exponential decay behavior (sometimes in absolute value depending on the signs of the AR and MA coefficients).

5.5.3 ACF and PACF of ARMA(*p, q*) Process

As in the stationarity and invertibility conditions, the ACF and PACF of an ARMA process are determined by the AR and MA components, respectively. It can therefore be shown that the ACF and PACF of an ARMA(p, q) both exhibit exponential decay and/or damped sinusoid patterns, which makes the identification of the order of the ARMA(p, q) model relatively more difficult. For that, additional sample functions such as the Extended Sample ACF (ESACF), the Generalized Sample PACF (GPACF), the Inverse ACF (IACF), and canonical correlations can be used. For further information see Box, Jenkins, and Reinsel (2008), Wei (2006), Tiao and Box (1981), Tsay and Tiao (1984), and Abraham and Ledolter (1984). However, the availability of sophisticated statistical software packages such as Minitab JMP and SAS makes it possible for the practitioner to consider several different models with various orders and compare them based on the model selection criteria such as AIC, AICC, and BIC as described in Chapter 2 and residual analysis.

The theoretical values of the ACF and PACF for stationary time series are summarized in Table 5.1. The summary of the sample ACFs and PACFs of the realizations of some of the models we have covered in this chapter are given in Table 5.2, Table 5.3, and Table 5.4 for MA, AR, and ARMA models, respectively.

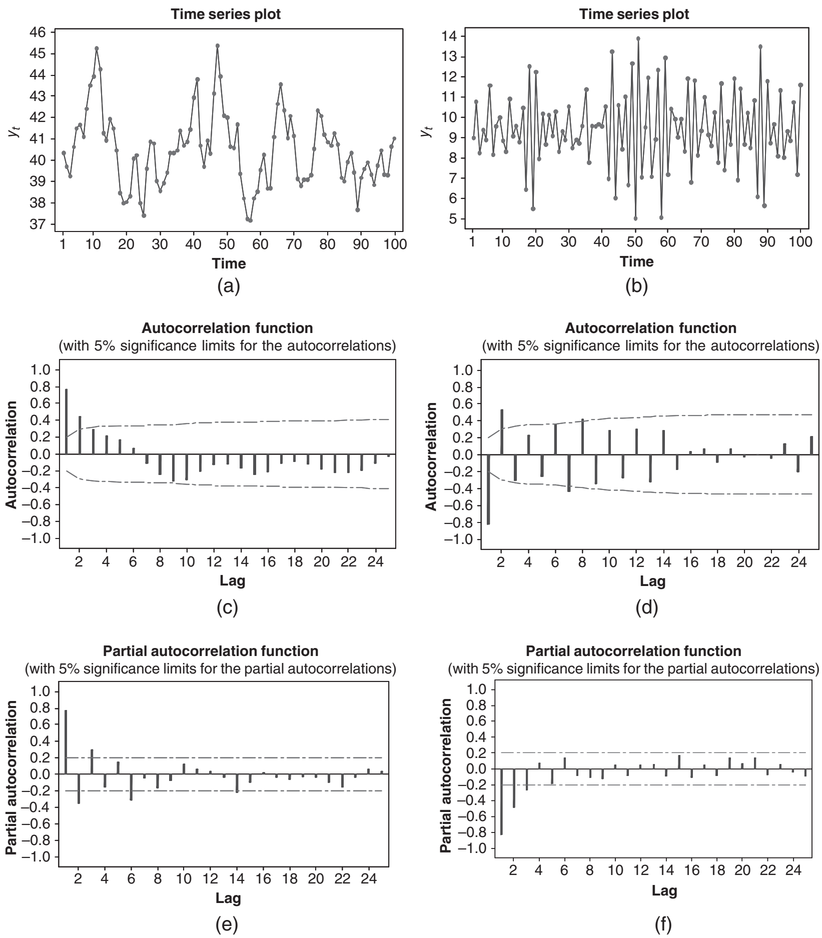
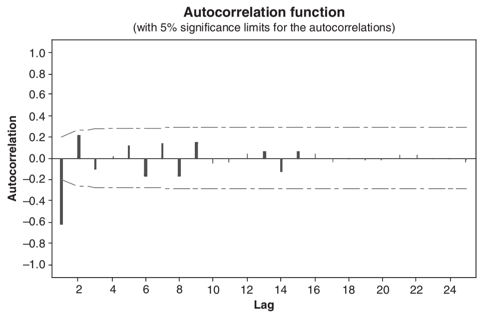
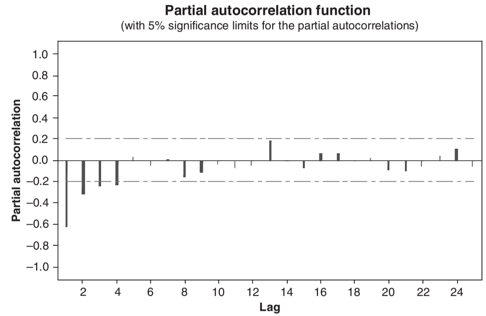
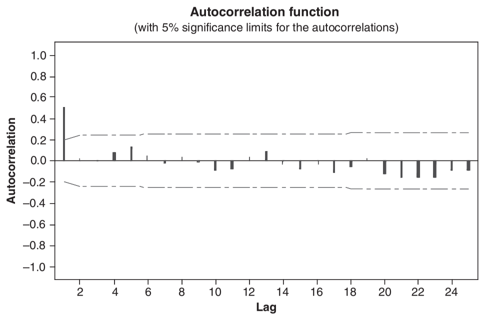
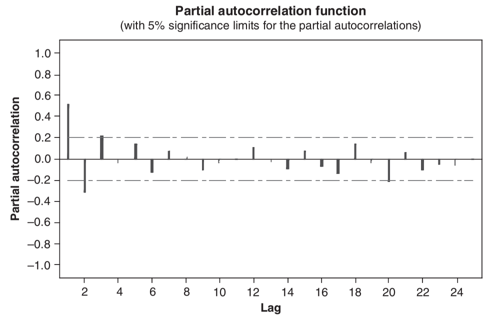


FIGURE 5.10 Two realizations of the ARMA(1,1) model: (a) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$ and (b) $y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$. (c) The ACF of (a), (d) the ACF of (b), (e) the PACF of (a), and (f) the PACF of (b).

TABLE 5.1 Behavior of Theoretical ACF and PACF for Stationary Processes

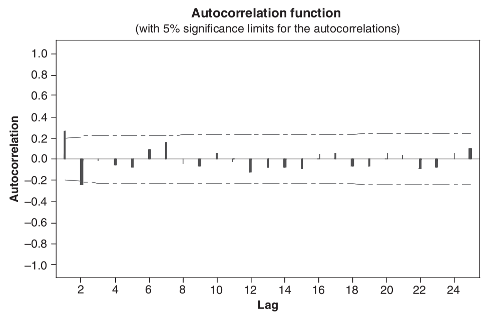
Model	ACF	PACF
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

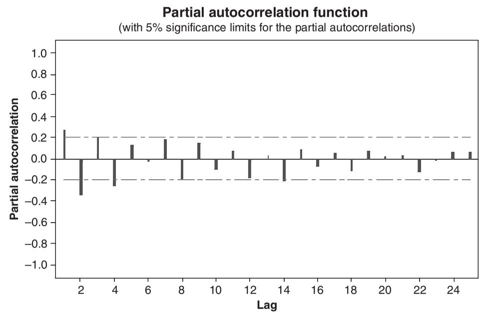
TABLE 5.2 Sample ACFs and PACFs for Some Realizations of MA(1) and MA(2) Models

Model	Sample ACF	Sample PACF
MA(1) $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$		
$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$		

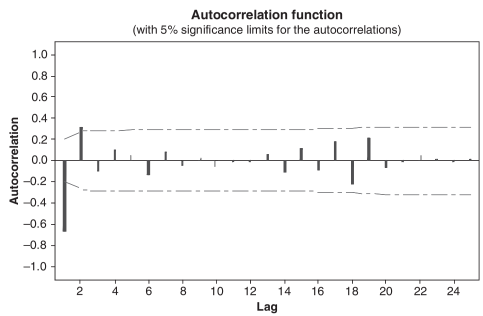
MA(2)

$$y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$$





$$y_t = 40 + \varepsilon_t - 1.1\varepsilon_{t-1} + 0.8\varepsilon_{t-2}$$



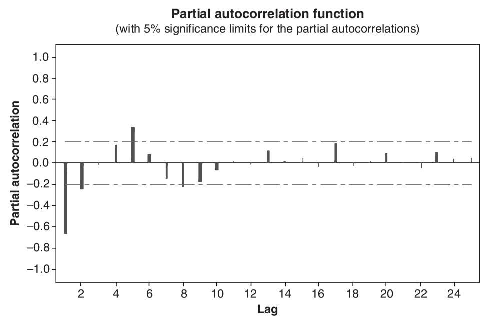
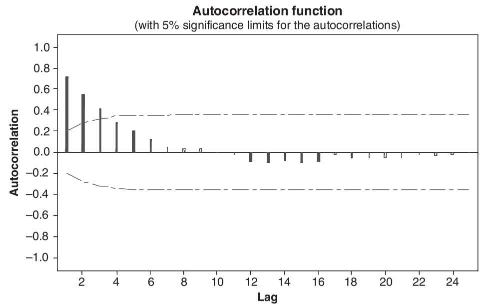
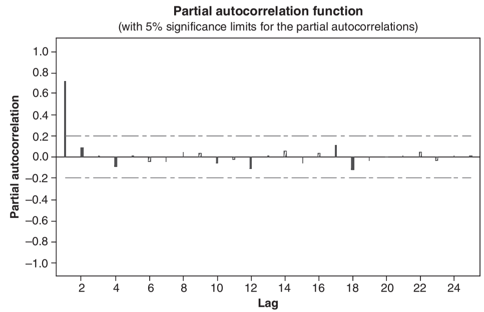
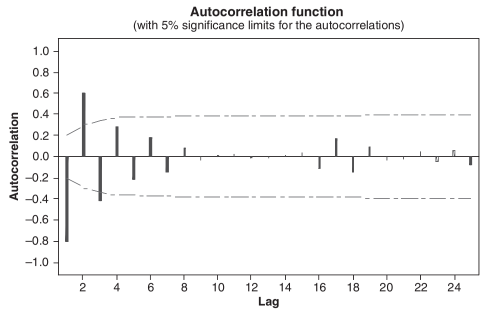
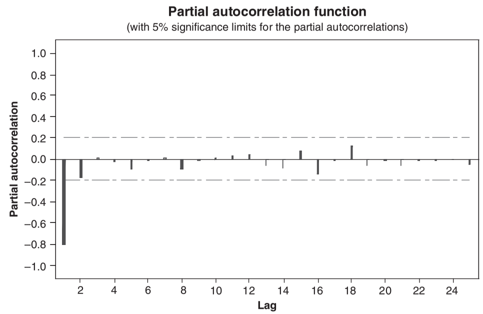
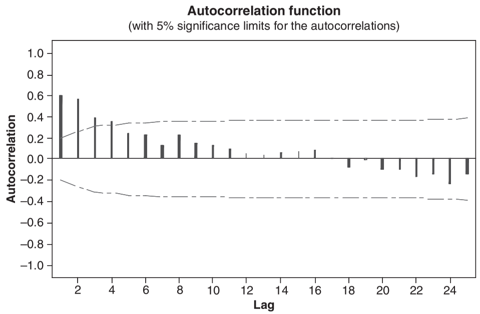


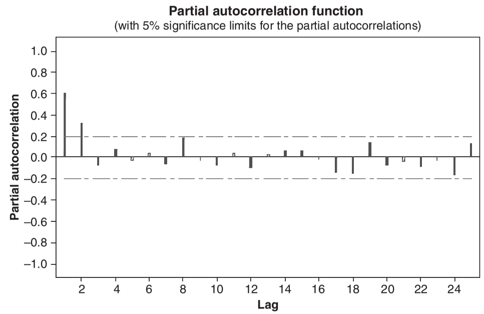
TABLE 5.3 Sample ACFs and PACFs for Some Realizations of AR(1) and AR(2) Models

Model	Sample ACF	Sample PACF
AR(1) $y_t = 8 + 0.8y_{t-1} + \varepsilon_t$		
$y_t = 8 - 0.8y_{t-1} + \varepsilon_t$		

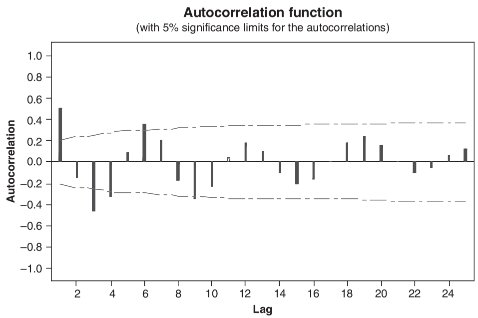
AR(2)

$$y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$$





$$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$$



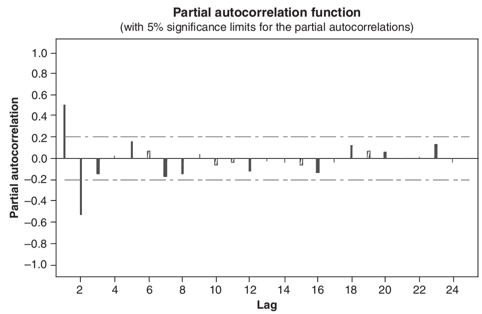
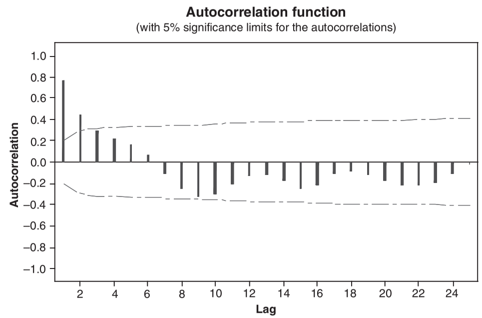
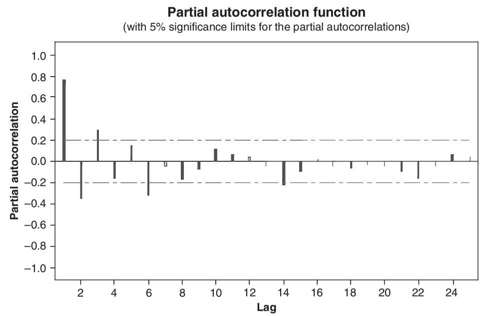
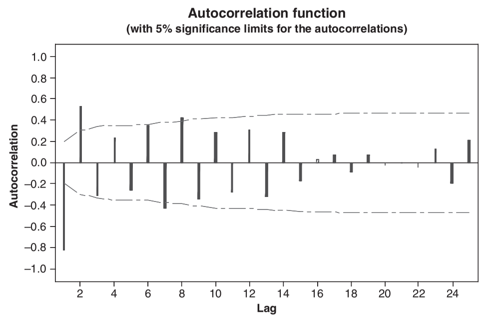
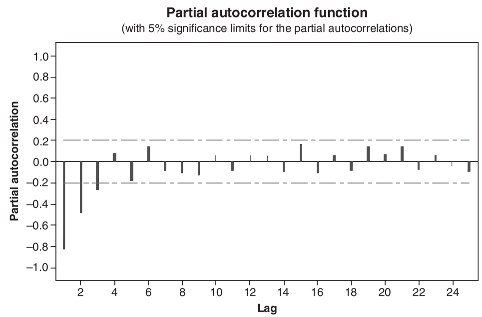


TABLE 5.4 Sample ACFs and PACFs for Some Realizations of ARMA(1,1) Models

Model	Sample ACF	Sample PACF
ARMA(1.1) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$		
$y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$		

5.6 NONSTATIONARY PROCESSES

It is often the case that while the processes may not have a constant level, they exhibit homogeneous behavior over time. Consider, for example, the linear trend process given in Figure 5.1c. It can be seen that different snapshots taken in time do exhibit similar behavior except for the mean level of the process. Similarly, processes may show nonstationarity in the slope as well. We will call a time series, y_t, homogeneous nonstationary if it is not stationary but its first difference, that is, $w_t = y_t - y_{t-1} = (1 - B)\, y_t$, or higher-order differences, $w_t = (1 - B)^d\, y_t$, produce a stationary time series. We will further call y_t an **autoregressive integrated moving average** (ARIMA) process of orders p, d, and q—that is, ARIMA(p, d, q)—if its dth difference, denoted by $w_t = (1 - B)^d\, y_t$, produces a stationary ARMA(p, q) process. The term integrated is used since, for $d = 1$, for example, we can write y_t as the sum (or "integral") of the w_t process as

$$\begin{aligned} y_t &= w_t + y_{t-1} \\ &= w_t + w_{t-1} + y_{t-2} \\ &= w_t + w_{t-1} + \cdots + w_1 + y_0 \end{aligned} \tag{5.76}$$

Hence an ARIMA(p, d, q) can be written as

$$\Phi(B)\,(1 - B)^d\, y_t = \delta + \Theta(B)\,\varepsilon_t \tag{5.77}$$

Thus once the differencing is performed and a stationary time series $w_t = (1 - B)^d\, y_t$ is obtained, the methods provided in the previous sections can be used to obtain the full model. In most applications first differencing ($d = 1$) and occasionally second differencing ($d = 2$) would be enough to achieve stationarity. However, sometimes transformations other than differencing are useful in reducing a nonstationary time series to a stationary one. For example, in many economic time series the variability of the observations increases as the average level of the process increases; however, the percentage of change in the observations is relatively independent of level. Therefore taking the logarithm of the original series will be useful in achieving stationarity.

5.6.1 Some Examples of ARIMA(*p, d, q*) Processes

The **random walk process, ARIMA(0, 1, 0)** is the simplest nonstationary model. It is given by

$$(1 - B)y_t = \delta + \varepsilon_t \tag{5.78}$$

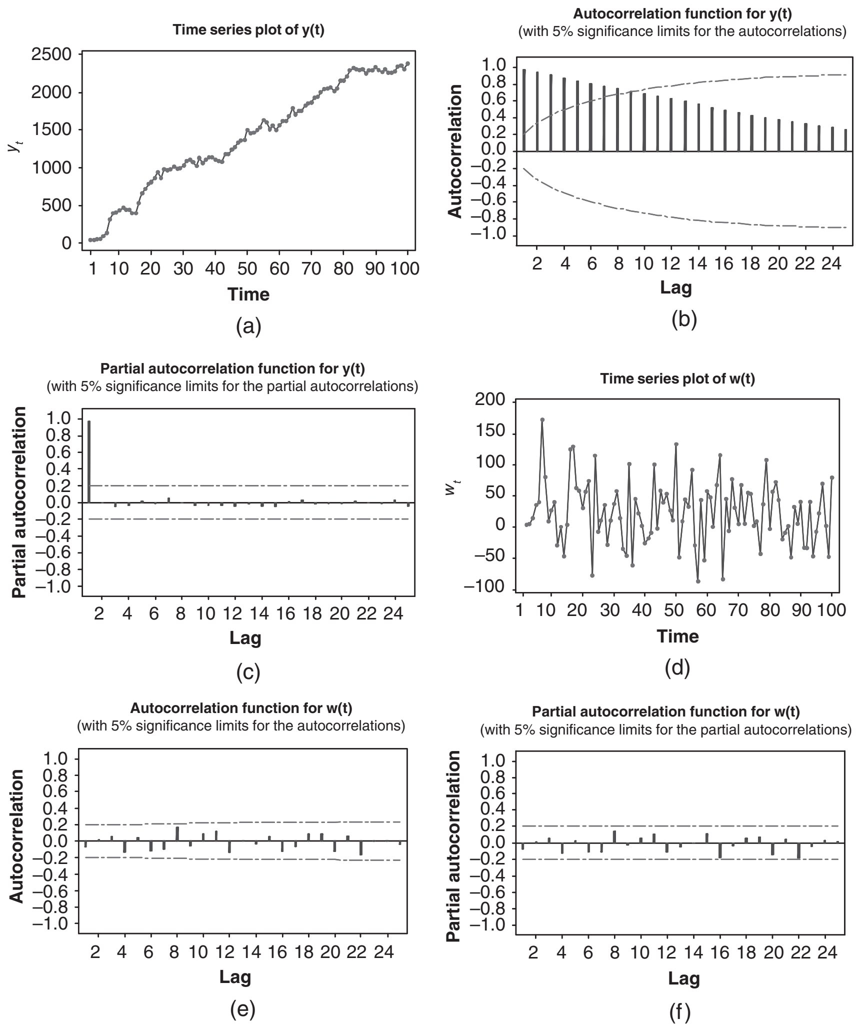


FIGURE 5.11 A realization of the ARIMA(0, 1, 0) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

suggesting that first differencing eliminates all serial dependence and yields a white noise process.

Consider the process $y_t = 20 + y_{t-1} + \varepsilon_t$. A realization of this process together with its sample ACF and PACF are given in Figure 5.11a–c. We can see that the sample ACF dies out very slowly, while the sample PACF is only significant at the first lag. Also note that the PACF value at the first lag is very close to one. All this evidence suggests that the process is not stationary. The first difference, $w_t = y_t - y_{t-1}$, and its sample ACF and PACF are shown in Figure 5.11d–f. The time series plot of w_t implies that the first difference is stationary. In fact, the sample ACF and PACF do not show any significant values. This further suggests that differencing the original data once "clears out" the autocorrelation. Hence the data can be modeled using the random walk model given in Eq. (5.78).

The **ARIMA(0, 1, 1) process** is given by

$$(1 - B)y_t = \delta + (1 - \theta B)\,\varepsilon_t \tag{5.79}$$

The infinite AR representation of Eq. (5.79) can be obtained from Eq. (5.75)

$$\pi_i - \theta\pi_{i-1} = \begin{cases} 1, & i = 1 \\ 0, & i > 1 \end{cases} \tag{5.80}$$

with $\pi_0 = -1$. Thus we have

$$\begin{aligned} y_t &= \alpha + \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t \\ &= \alpha + (1 - \theta)(y_{t-1} + \theta y_{t-2} + \cdots) + \varepsilon_t \end{aligned} \tag{5.81}$$

This suggests that an ARIMA(0, 1, 1) (a.k.a. IMA(1, 1)) can be written as an exponentially weighted moving average (EWMA) of all past values.

Consider the time series data in Figure 5.12a. It looks like the mean of the process is changing (moving upwards) in time. Yet the change in the mean (i.e., nonstationarity) is not as obvious as in the previous example. The sample ACF plot of the data in Figure 5.12b dies down relatively slowly and the sample PACF of the data in Figure 5.12c shows two significant values at lags 1 and 2. Hence we might be tempted to model this data using an AR(2) model because of the exponentially decaying ACF and significant PACF at the first two lags. Indeed, we might even have a good fit using an AR(2) model. We should nevertheless check the roots of the associated polynomial given in Eq. (5.36) to make sure that its roots are less than 1 in absolute value. Also note that a technically stationary process will behave more and more nonstationary as the roots of the associated polynomial approach unity. For that, observe the realization of the near nonstationary process, $y_t = 2 + 0.95y_{t-1} + \varepsilon_t$, given in Figure 5.1b. Based on the visual inspection, however, we may deem the process nonstationary and proceed with taking the first difference of the data. This is because the ϕ value of the AR(1) model is close to 1. Under these circumstances, where the nonstationarity

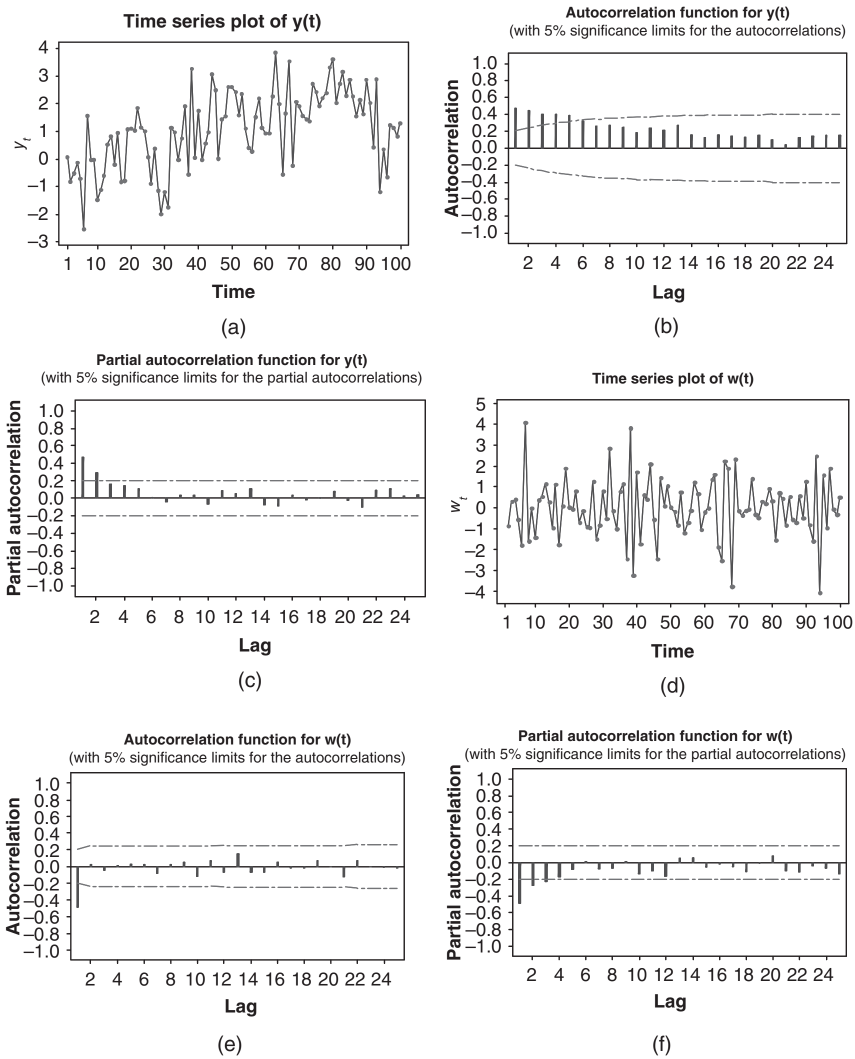


FIGURE 5.12 A realization of the ARIMA(0, 1, 1) model, y_t, its first difference, w_t, and their sample ACFs and PACFs.

of the process is dubious, we strongly recommend that the analyst refer back to basic underlying process knowledge. If, for example, the process mean is expected to wander off as in some financial data, assuming that the process is nonstationary and proceeding with differencing the data would be more appropriate. For the data given in Figure 5.12a, its first difference given in Figure 5.12d looks stationary. Furthermore, its sample ACF and

PACF given in Figures 5.12e and 5.12f, respectively, suggest that an MA(1) model would be appropriate for the first difference since its ACF cuts off after the first lag and the PACF exhibits an exponential decay pattern. Hence the ARIMA(0, 1, 1) model given in Eq. (5.79) can be used for this data.

5.7 TIME SERIES MODEL BUILDING

A three-step iterative procedure is used to build an ARIMA model. First, a tentative model of the ARIMA class is identified through analysis of historical data. Second, the unknown parameters of the model are estimated. Third, through residual analysis, diagnostic checks are performed to determine the adequacy of the model, or to indicate potential improvements. We shall now discuss each of these steps in more detail.

5.7.1 Model Identification

Model identification efforts should start with preliminary efforts in understanding the type of process from which the data is coming and how it is collected. The process' perceived characteristics and sampling frequency often provide valuable information in this preliminary stage of model identification. In today's data rich environments, it is often expected that the practitioners would be presented with "enough" data to be able to generate reliable models. It would nevertheless be recommended that 50 or preferably more observations should be initially considered. Before engaging in rigorous statistical model-building efforts, we also strongly recommend the use of "creative" plotting of the data, such as the simple time series plot and scatter plots of the time series data y_t versus y_{t-1}, y_{t-2}, and so on. For the y_t versus y_{t-1} scatter plot, for example, this can be achieved in a data set of N observations by plotting the first $N - 1$ observations versus the last $N - 1$. Simple time series plots should be used as the preliminary assessment tool for stationarity. The visual inspection of these plots should later be confirmed as described earlier in this chapter. If nonstationarity is suspected, the time series plot of the first (or dth) difference should also be considered. The unit root test by Dickey and Fuller (1979) can also be performed to make sure that the differencing is indeed needed. Once the stationarity of the time series can be presumed, the sample ACF and PACF of the time series of the original time series (or its dth difference if necessary) should be obtained. Depending on the nature of the autocorrelation, the first 20–25 sample autocorrelations and partial autocorrelations should be sufficient. More care should be taken of course if the process exhibits strong autocorrelation and/or seasonality, as we will discuss in the following sections. Table 5.1 together with the $\pm 2/\sqrt{N}$ limits can be used as a guide for identifying AR or MA models. As discussed earlier, the identification of ARMA models would require more care, as both the ACF and PACF will exhibit exponential decay and/or damped sinusoid behavior.

We have already discussed that the differenced series $\{w_t\}$ may have a nonzero mean, say, μ_w. At the identification stage we may obtain an indication of whether or not a nonzero value of μ_w is needed by comparing the sample mean of the differenced series, say, $\bar{w} = \sum_{t=1}^{n-d} [w/(n-d)]$, with its approximate standard error. Box, Jenkins, and Reinsel (2008) give the approximate standard error of $\bar{w}$ for several useful ARIMA(p, d, q) models.

Identification of the appropriate ARIMA model requires skills obtained by experience. Several excellent examples of the identification process are given in Box et al. (2008, Chap. 6), Montgomery et al. (1990), and Bisgaard and Kulahci (2011).

5.7.2 Parameter Estimation

There are several methods such as the methods of moments, maximum likelihood, and least squares that can be employed to estimate the parameters in the tentatively identified model. However, unlike the regression models of Chapter 2, most ARIMA models are **nonlinear** models and require the use of a nonlinear model fitting procedure. This is usually automatically performed by sophisticated software packages such as Minitab JMP, and SAS. In some software packages, the user may have the choice of estimation method and can accordingly choose the most appropriate method based on the problem specifications.

5.7.3 Diagnostic Checking

After a tentative model has been fit to the data, we must examine its adequacy and, if necessary, suggest potential improvements. This is done through residual analysis. The residuals for an ARMA(p, q) process can be obtained from

$$\hat{\varepsilon}_t = y_t - \left(\hat{\delta} + \sum_{i=1}^{p} \hat{\phi}_i y_{t-i} - \sum_{i=1}^{q} \hat{\theta}_i \hat{\varepsilon}_{t-i} \right) \tag{5.82}$$

If the specified model is adequate and hence the appropriate orders p and q are identified, it should transform the observations to a white noise process. Thus the residuals in Eq. (5.82) should behave like white noise.

Let the sample autocorrelation function of the residuals be denoted by $\{r_e(k)\}$. If the model is appropriate, then the residual sample autocorrelation function should have no structure to identify. That is, the autocorrelation should not differ significantly from zero for all lags greater than one. If the form of the model were correct and if we knew the true parameter values, then the standard error of the residual autocorrelations would be $N^{-1/2}$.

Rather than considering the $r_e(k)$ terms individually, we may obtain an indication of whether the first K residual autocorrelations considered together indicate adequacy of the model. This indication may be obtained through an approximate chi-square test of model adequacy. The test statistic is

$$Q = (N - d) \sum_{k=1}^{K} r_e^2(k) \tag{5.83}$$

which is approximately distributed as chi-square with $K - p - q$ degrees of freedom if the model is appropriate. If the model is inadequate, the calculated value of Q will be too large. Thus we should reject the hypothesis of model adequacy if Q exceeds an approximate small upper tail point of the chi-square distribution with $K - p - q$ degrees of freedom. Further details of this test are in Chapter 2 and in the original reference by Box and Pierce (1970). The modification of this test by Ljung and Box (1978) presented in Chapter 2 is also useful in assessing model adequacy.

5.7.4 Examples of Building ARIMA Models

In this section we shall present two examples of the identification, estimation, and diagnostic checking process. One example presents the analysis for a stationary time series, while the other is an example of modeling a nonstationary series.

Example 5.1 Table 5.5 shows the weekly total number of loan applications in a local branch of a national bank for the last 2 years. It is suspected that there should be some relationship (i.e., autocorrelation) between the number of applications in the current week and the number of loan applications in the previous weeks. Modeling that relationship will help the management to proactively plan for the coming weeks through reliable forecasts. As always, we start our analysis with the time series plot of the data, shown in Figure 5.13.

TABLE 5.5 Weekly Total Number of Loan Applications for the Last 2 Years

Week	Applications	Week	Applications	Week	Applications	Week	Applications
1	71	27	62	53	66	79	63
2	57	28	77	54	71	80	61
3	62	29	76	55	59	81	73
4	64	30	88	56	57	82	72
5	65	31	71	57	66	83	65
6	67	32	72	58	51	84	70
7	65	33	66	59	59	85	54
8	82	34	65	60	56	86	63
9	70	35	73	61	57	87	62
10	74	36	76	62	55	88	60
11	75	37	81	63	53	89	67
12	81	38	84	64	74	90	59
13	71	39	68	65	64	91	74
14	75	40	63	66	70	92	61
15	82	41	66	67	74	93	61
16	74	42	71	68	69	94	52
17	78	43	67	69	64	95	55
18	75	44	69	70	68	96	61
19	73	45	63	71	64	97	56
20	76	46	61	72	70	98	61
21	66	47	68	73	73	99	60
22	69	48	75	74	59	100	65
23	63	49	66	75	68	101	55
24	76	50	81	76	59	102	61
25	65	51	72	77	66	103	59
26	73	52	77	78	63	104	63

Figure 5.13 shows that the weekly data tend to have short runs and that the data seem to be indeed autocorrelated. Next, we visually inspect the stationarity. Although there might be a slight drop in the mean for the second year (weeks 53–104), in general, it seems to be safe to assume stationarity.

We now look at the sample ACF and PACF plots in Figure 5.14. Here are possible interpretations of the ACF plot:

1. It cuts off after lag 2 (or maybe even 3), suggesting an MA(2) (or MA(3)) model.
2. It has an (or a mixture of) exponential decay(s) pattern suggesting an AR(p) model.

To resolve the conflict, consider the sample PACF plot. For that, we have only one interpretation; it cuts off after lag 2. Hence we use the second

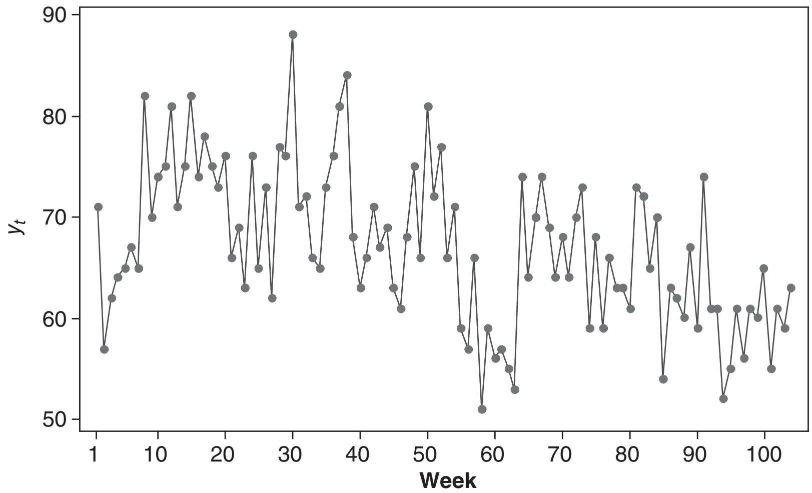


FIGURE 5.13 Time series plot of the weekly total number of loan applications.

interpretation of the sample ACF plot and assume that the appropriate model to fit is the AR(2) model.

Table 5.6 shows the Minitab output for the AR(2) model. The parameter estimates are $\hat{\phi}_1 = 0.27$ and $\hat{\phi}_2 = 0.42$, and they turn out to be significant (see the P-values).

MSE is calculated to be 39.35. The modified Box–Pierce test suggests that there is no autocorrelation left in the residuals. We can also see this in the ACF and PACF plots of the residuals in Figure 5.15.

As the last diagnostic check, we have the 4-in-1 residual plots in Figure 5.16 provided by Minitab: Normal Probability Plot, Residuals versus

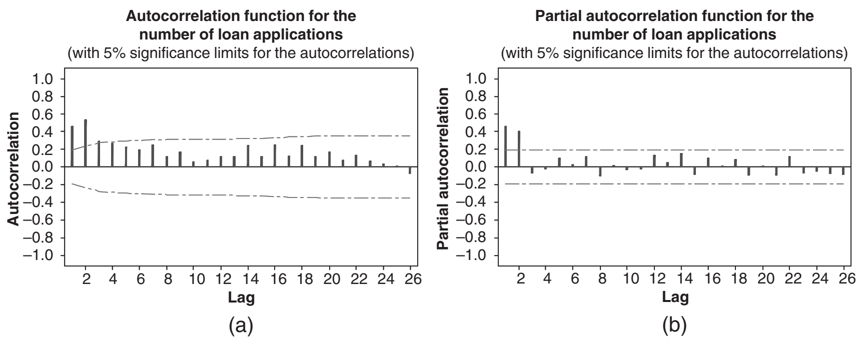


FIGURE 5.14 ACF and PACF for the weekly total number of loan applications.

TABLE 5.6 Minitab Output for the AR(2) Model for the Loan Application Data

```
Final Estimates of Parameters

Type          Coef  SE Coef       T      P
AR   1      0.2682   0.0903    2.97  0.004
AR   2      0.4212   0.0908    4.64  0.000
Constant   20.7642   0.6157   33.73  0.000
Mean        66.844    1.982


Number of observations:  104
Residuals:    SS =  3974.30 (backforecasts excluded)
              MS =  39.35  DF = 101


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag            12     24     36     48
Chi-Square    6.2   16.0   24.9   32.0
DF              9     21     33     45
P-Value     0.718  0.772  0.843  0.927
```

Fitted Value, Histogram of the Residuals, and Time Series Plot of the Residuals. They indicate that the fit is indeed acceptable.

Figure 5.17 shows the actual data and the fitted values. It looks like the fitted values smooth out the highs and lows in the data.

Note that, in this example, we often and deliberately used "vague" words such as "seems" or "looks like." It should be clear by now that

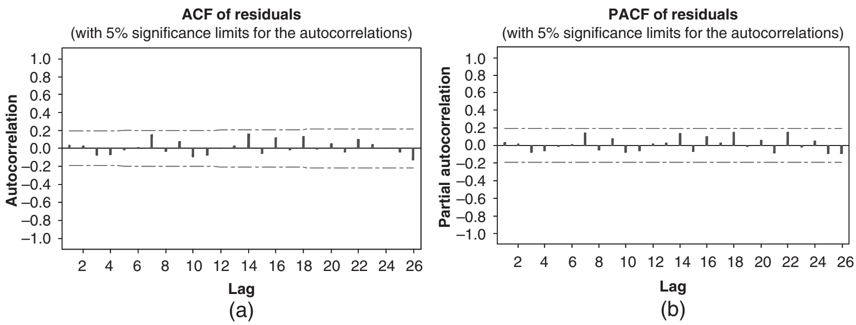


FIGURE 5.15 The sample ACF and PACF of the residuals for the AR(2) model in Table 5.6.

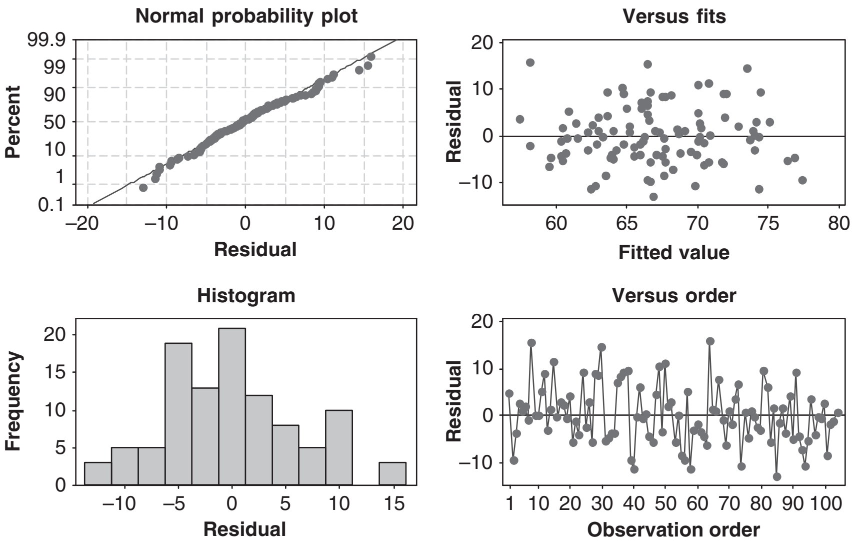


FIGURE 5.16 Residual plots for the AR(2) model in Table 5.6.

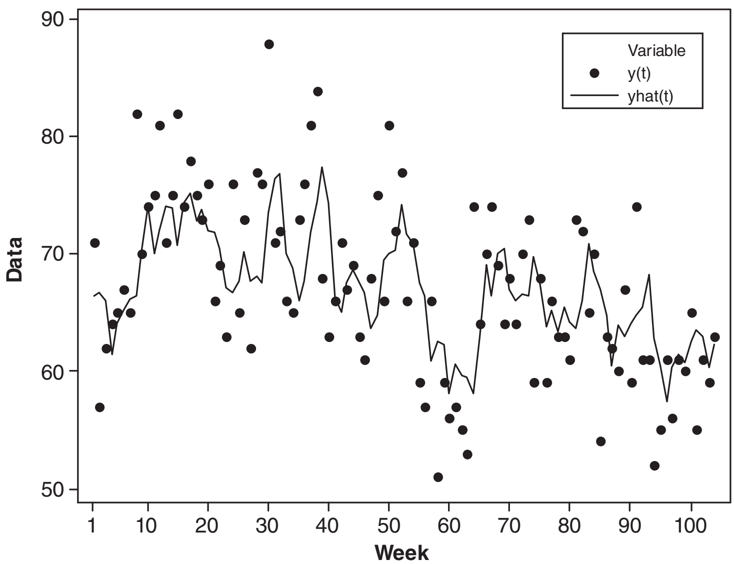


FIGURE 5.17 Time series plot of the actual data and fitted values for the AR(2) model in Table 5.6.

the methodology presented in this chapter has a very sound theoretical foundation. However, as in any modeling effort, we should also keep in mind the subjective component of model identification. In fact, as we mentioned earlier, time series model fitting can be seen as a mixture of science and art and can best be learned by practice and experience. The next example will illustrate this point further.

Example 5.2 Consider the Dow Jones Index data from Chapter 4. A time series plot of the data is given in Figure 5.18. The process shows signs of nonstationarity with changing mean and possibly variance.

Similarly, the slowly decreasing sample ACF and sample PACF with significant value at lag 1, which is close to 1 in Figure 5.19, confirm that indeed the process can be deemed nonstationary. On the other hand, one might argue that the significant sample PACF value at lag 1 suggests that the AR(1) model might also fit the data well. We will consider this interpretation first and fit an AR(1) model to the Dow Jones Index data.

Table 5.7 shows the Minitab output for the AR(1) model. Although it is close to 1, the AR(1) model coefficient estimate $\hat{\phi} = 0.9045$ turns out to be quite significant and the modified Box–Pierce test suggests that there is no autocorrelation left in the residuals. This is also confirmed by the sample ACF and PACF plots of the residuals given in Figure 5.20.

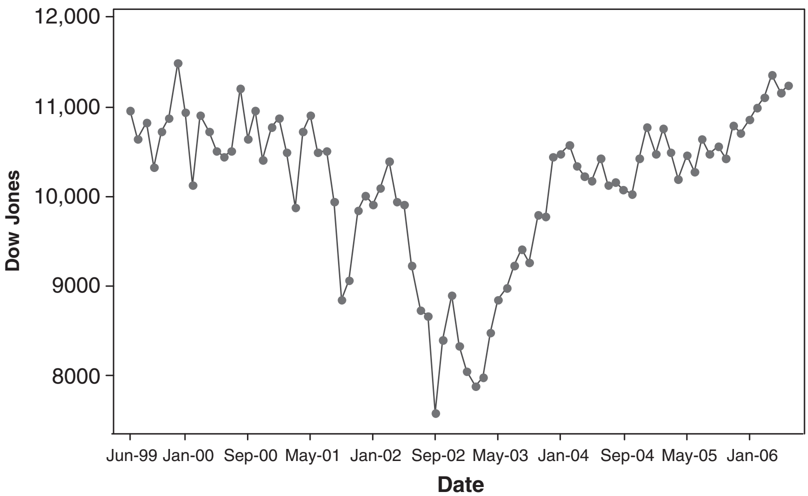


FIGURE 5.18 Time series plot of the Dow Jones Index from June 1999 to June 2006.

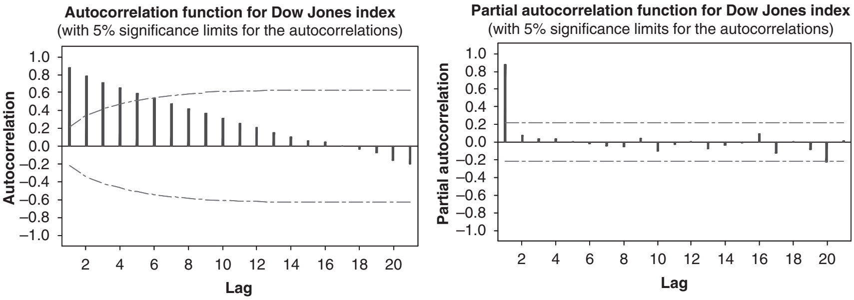


FIGURE 5.19 Sample ACF and PACF of the Dow Jones Index.

TABLE 5.7 Minitab Output for the AR(1) Model for the Dow Jones Index

```
Final Estimates of Parameters

Type           Coef   SE Coef       T       P
AR   1       0.9045    0.0500   18.10   0.000
Constant     984.94     44.27   22.25   0.000
Mean        10309.9     463.4


Number of observations:  85
Residuals:     SS =  13246015 (backforecasts excluded)
               MS =  159591  DF = 83


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag              12      24      36      48
Chi-Square      2.5    14.8    21.4    29.0
DF               10      22      34      46
P-Value       0.991   0.872   0.954   0.977
```

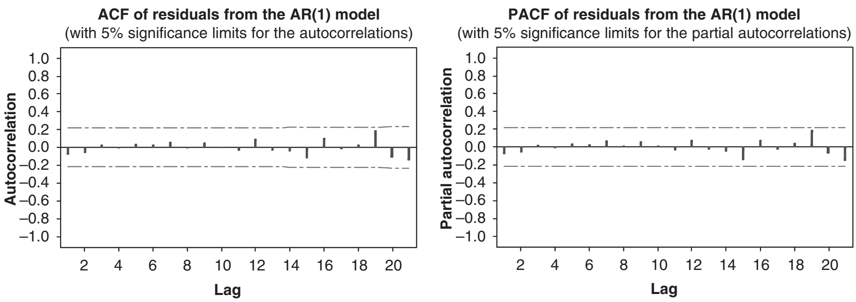


FIGURE 5.20 Sample ACF and PACF of the residuals from the AR(1) model for the Dow Jones Index data.

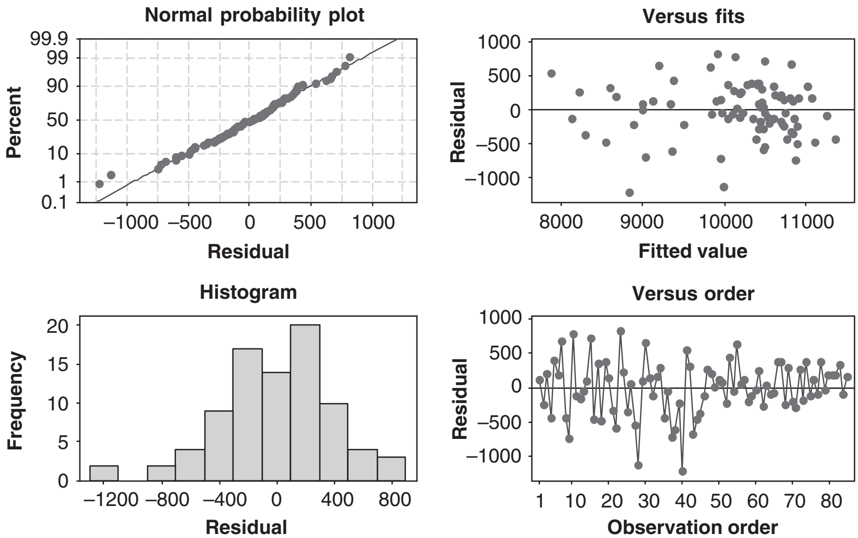


FIGURE 5.21 Residual plots from the AR(1) model for the Dow Jones Index data.

The only concern in the residual plots in Figure 5.21 is in the changing variance observed in the time series plot of the residuals. This is indeed a very important issue since it violates the constant variance assumption. We will discuss this issue further in Section 7.3 but for illustration purposes we will ignore it in this example.

Overall it can be argued that an AR(1) model provides a decent fit to the data. However, we will now consider the earlier interpretation and assume that the Dow Jones Index data comes from a nonstationary process. We then take the first difference of the data as shown in Figure 5.22. While there are once again some serious concerns about changing variance, the level of the first difference remains the same. If we ignore the changing variance and look at the sample ACF and PACF plots given in Figure 5.23, we may conclude that the first difference is in fact white noise. That is, since these plots do not show any sign of significant autocorrelation, a model we may consider for the Dow Jones Index data would be the random walk model, ARIMA(0, 1, 0).

Now the analyst has to decide between the two models: AR(1) and ARIMA(0, 1, 0). One can certainly use some of the criteria we discussed in Section 2.6.2 to choose one of these models. Since these two models are fundamentally quite different, we strongly recommend that the analyst use the subject matter/process knowledge as much as possible. Do we expect

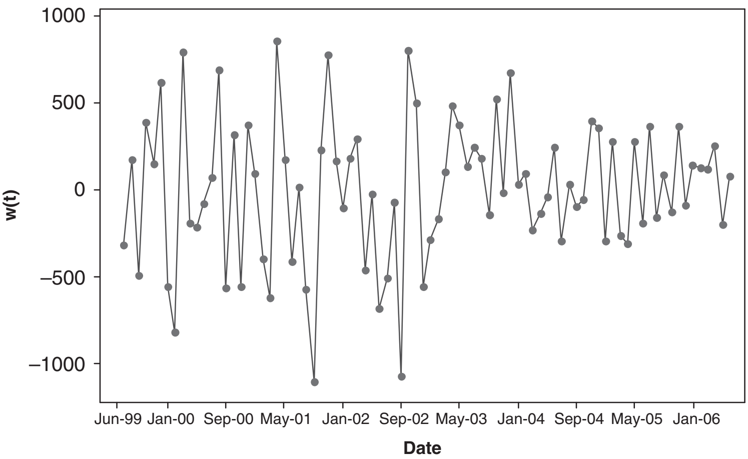


FIGURE 5.22 Time series plot of the first difference $w(t)$ of the Dow Jones Index data.

a financial index such as the Dow Jones Index to wander about a fixed mean as implied by the AR(1)? In most cases involving financial data, the answer would be no. Hence a model such as ARIMA(0, 1, 0) that takes into account the inherent nonstationarity of the process should be preferred. However, we do have a problem with the proposed model. A random walk model means that the price changes are random and cannot be predicted. If we have a higher price today compared to yesterday, that would have no bearing on the forecasts tomorrow. That is, tomorrow's price can be higher or lower than today's and we would have no way to forecast it effectively. This further suggests that the best forecast for tomorrow's price is in fact the price we have today. This is obviously not a reliable and effective forecasting model. This very same issue of the random walk models for

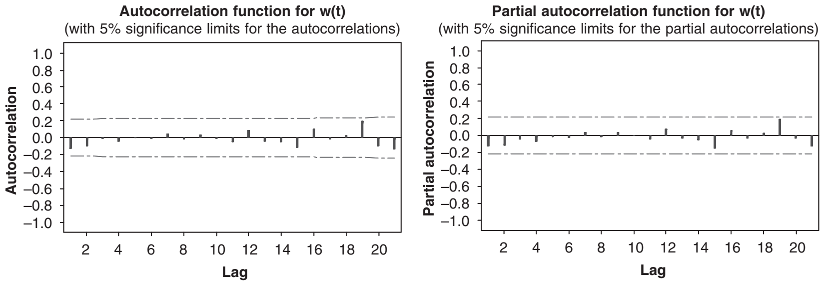


FIGURE 5.23 Sample ACF and PACF plots of the first difference of the Dow Jones Index data.

financial data has been discussed in great detail in the literature. We simply used this data to illustrate that in time series model fitting we can end up with fundamentally different models that will fit the data equally well. At this point, process knowledge can provide the needed guidance in picking the "right" model.

It should be noted that, in this example, we tried to keep the models simple for illustration purposes. Indeed, a more thorough analysis would (and should) pay close attention to the changing variance issue. In fact, this is a very common concern particularly when dealing with financial data. For that, we once again refer the reader to Section 7.3.

5.8 FORECASTING ARIMA PROCESSES

Once an appropriate time series model has been fit, it may be used to generate forecasts of future observations. If we denote the current time by T, the forecast for $y_{T+\tau}$ is called the τ-period-ahead forecast and denoted by $\hat{y}_{T+\tau}(T)$. The standard criterion to use in obtaining the best forecast is the mean squared error for which the expected value of the squared forecast errors, $E[(y_{T+\tau} - \hat{y}_{T+\tau}(T))^2] = E[e_T(\tau)^2]$, is minimized. It can be shown that the best forecast in the mean square sense is the conditional expectation of $y_{T+\tau}$ given current and previous observations, that is, y_T, $y_{T-1}, \ldots$:

$$\hat{y}_{T+\tau}(T) = E\left[y_{T+\tau} \middle| y_T, y_{T-1}, \ldots\right] \tag{5.84}$$

Consider, for example, an ARIMA(p, d, q) process at time $T+\tau$ (i.e., τ period in the future):

$$y_{T+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{T+\tau-i} + \varepsilon_{T+\tau} - \sum_{i=1}^{q} \theta_i \varepsilon_{T+\tau-i} \tag{5.85}$$

Further consider its infinite MA representation,

$$y_{T+\tau} = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.86}$$

We can partition Eq. (5.86) as

$$y_{T+\tau} = \mu + \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.87}$$

In this partition, we can clearly see that the $\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}$ component involves the future errors, whereas the $\sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i}$ component involves the present and past errors. From the relationship between the current and past observations and the corresponding random shocks as well as the fact that the random shocks are assumed to have mean zero and to be independent, we can show that the best forecast in the mean square sense is

$$\hat{y}_{T+\tau}(T) = E\left[y_{T+\tau} \middle| y_T, y_{T-1}, \ldots\right] = \mu + \sum_{i=\tau}^{\infty} \psi_i \varepsilon_{T+\tau-i} \tag{5.88}$$

since

$$E\left[\varepsilon_{T+\tau-i} \middle| y_T, y_{T-1}, \ldots\right] = \begin{cases} 0 & \text{if} \quad i < \tau \\ \varepsilon_{T+\tau-i} & \text{if} \quad i \geq \tau \end{cases}$$

Subsequently, the forecast error is calculated from

$$e_T(\tau) = y_{T+\tau} - \hat{y}_{T+\tau}(T) = \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i} \tag{5.89}$$

Since the forecast error in Eq. (5.89) is a linear combination of random shocks, we have

$$E\left[e_T(\tau)\right] = 0 \tag{5.90}$$

$$\begin{aligned} \operatorname{Var}\left[e_T(\tau)\right] &= \operatorname{Var}\left[\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{T+\tau-i}\right] = \sum_{i=0}^{\tau-1} \psi_i^2 \operatorname{Var}(\varepsilon_{T+\tau-i}) \\ &= \sigma^2 \sum_{i=0}^{\tau-1} \psi_i^2 \\ &= \sigma^2(\tau), \quad \tau = 1, 2, \ldots \end{aligned} \tag{5.91}$$

It should be noted that the variance of the forecast error gets bigger with increasing forecast lead times τ. This intuitively makes sense as we should expect more uncertainty in our forecasts further into the future. Moreover, if the random shocks are assumed to be normally distributed, $N(0, \sigma^2)$, then the forecast errors will also be normally distributed with $N(0, \sigma^2(\tau))$. We can then obtain the $100(1-\alpha)$ percent prediction intervals for the future observations from

$$P\left(\hat{y}_{T+\tau}(T) - z_{\alpha/2}\sigma(\tau) < y_{T+\tau} < \hat{y}_{T+\tau}(T) + z_{\alpha/2}\sigma(\tau)\right) = 1-\alpha \tag{5.92}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ percentile of the standard normal distribution, $N(0,1)$. Hence the $100(1-\alpha)$ percent prediction interval for $y_{T+\tau}$ is

$$\hat{y}_{T+\tau}(T) \pm z_{\alpha/2}\sigma(\tau) \tag{5.93}$$

There are two issues with the forecast equation in (5.88). First, it involves infinitely many terms in the past. However, in practice, we will only have a finite amount of data. For a sufficiently large data set, this can be overlooked. Second, Eq. (5.88) requires knowledge of the magnitude of random shocks in the past, which is unrealistic. A solution to this problem is to "estimate" the past random shocks through one-step-ahead forecasts. For the ARIMA model we can calculate

$$\hat{\varepsilon}_t = y_t - \left[\delta + \sum_{i=1}^{p+d} \phi_i y_{t-i} - \sum_{i=1}^{q} \theta_i \hat{\varepsilon}_{t-i}\right] \tag{5.94}$$

recursively by setting the initial values of the random shocks to zero for $t < p+d+1$. For more accurate results, these initial values together with the y_t for $t \le 0$ can also be obtained using back-forecasting. For further details, see Box, Jenkins, and Reinsel (2008).

As an illustration consider forecasting the ARIMA(1, 1, 1) process

$$(1-\phi B)(1-B)\,y_{T+\tau} = (1-\theta B)\,\varepsilon_{T+\tau} \tag{5.95}$$

We will consider two of the most commonly used approaches:

1. As discussed earlier, this approach involves the infinite MA representation of the model in Eq. (5.95), also known as the **random shock** form of the model:

$$\begin{aligned} y_{T+\tau} &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{T+\tau-i} \\ &= \psi_0 \varepsilon_{T+\tau} + \psi_1 \varepsilon_{T+\tau-1} + \psi_2 \varepsilon_{T+\tau-2} + \cdots \end{aligned} \tag{5.96}$$

Hence the τ-step-ahead forecast can be calculated from

$$\hat{y}_{T+\tau}(T) = \psi_\tau \varepsilon_T + \psi_{\tau+1}\varepsilon_{T-1} + \cdots \tag{5.97}$$

The weights ψ_i can be calculated from

$$\left(\psi_0 + \psi_1 B + \cdots\right)(1 - \phi B)(1 - B) = (1 - \theta B) \tag{5.98}$$

and the random shocks can be estimated using the one-step-ahead forecast error; for example, ε_T can be replaced by $e_{T-1}(1) = y_T - \hat{y}_T(T-1)$.

2. Another approach that is often employed in practice is to use **difference equations** as given by

$$y_{T+\tau} = (1+\phi)\, y_{T+\tau-1} - \phi y_{T+\tau-2} + \varepsilon_{T+\tau} - \theta \varepsilon_{T+\tau-1} \tag{5.99}$$

For $\tau = 1$, the best forecast in the mean squared error sense is

$$\hat{y}_{T+1}(T) = E\left[y_{T+1} \middle| y_T, y_{T-1}, \ldots\right] = (1+\phi)\, y_T - \phi y_{T-1} - \theta e_T(1) \tag{5.100}$$

We can further show that for lead times $\tau > 2$, the forecast is

$$\hat{y}_{T+\tau}(T) = (1-\phi)\, \hat{y}_{T+\tau-1}(T) - \phi \hat{y}_{T+\tau-2}(T) \tag{5.101}$$

Prediction intervals for forecasts of future observations at time period $T + \tau$ are found using equation 5.87. However, in using Equation 5.87 the ψ weights must be found in order to compute the variance (or standard deviation) of the τ-step ahead forecast error. The ψ weights for the general ARIMA(p, d, q) model may be obtained by equating like powers of B in the expansion of

$$\begin{aligned}&(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1 - B)^d \\ &\quad = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\end{aligned}$$

and solving for the ψ weights. We now illustrate this with three examples.

Example 5.3 The ARMA(1, 1) Model For the ARMA(1, 1) model the product of the required polynomials is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi B) = (1 - \theta B)$$

Equating like power of B we find that

$$
\begin{aligned}
&B^0\colon \psi_0 = 1\\
&B^1\colon \psi_1 - \phi = -\theta, \text{ or } \psi_1 = \phi - \theta\\
&B^2\colon \psi_2 - \phi\psi_1 = 0, \text{ or } \psi_2 = \phi(\phi - \theta)
\end{aligned}
$$

In general, we can show for the ARMA(1,1) model that $\psi_j = \phi^{j-1}(\phi - \theta)$.

Example 5.4 The AR(2) Model For the AR(2) model the product of the required polynomials is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - \phi_1 B - \phi_2 B^2) = 1$$

Equating like power of B, we find that

$$
\begin{aligned}
&B^0\colon \psi_0 = 1\\
&B^1\colon \psi_1 - \phi_1 = 0, \text{ or } \psi_1 = \phi_1\\
&B^2\colon \psi_2 - \phi_1\psi_1 - \phi_2 = 0, \text{ or } \psi_2 = \phi_1\psi_1 + \phi_2
\end{aligned}
$$

In general, we can show for the AR(2) model that $\psi_j = \phi_1\psi_{j-1} + \phi_2\psi_{j-2}$.

Example 5.5 The ARIMA(0, 1, 1) or IMA(1,1) Model Now consider a nonstationary model, the IMA(1, 1) model. The product of the required polynomials for this model is

$$(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots)(1 - B) = (1 - \theta B)$$

It is straightforward to show that the ψ weights for this model are

$$
\begin{aligned}
&\psi_0 = 1\\
&\psi_1 = 1 - \theta\\
&\psi_j = \psi_{j-1},\ j = 2, 3, \ldots
\end{aligned}
$$

Notice that the prediction intervals will increase in length rapidly as the forecast lead time increases. This is typical of nonstationary ARIMA models. It implies that these models may not be very effective in forecasting more than a few periods ahead.

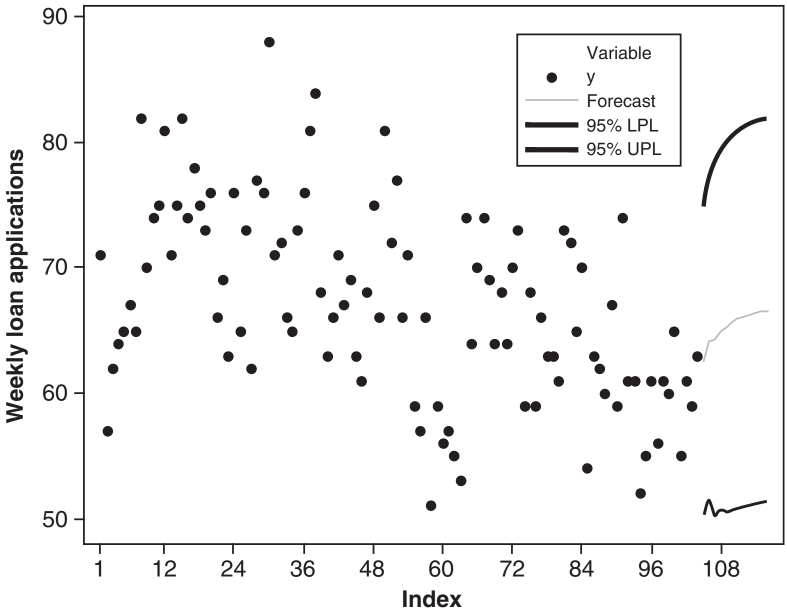


FIGURE 5.24 Time series plot and forecasts for the weekly loan application data.

Example 5.6 Consider the loan applications data given in Table 5.5. Now assume that the manager wants to make forecasts for the next 3 months (12 weeks) using the AR(2) model from Example 5.1. Hence at the 104th week we need to make 1-step, 2-step, … , 12-step-ahead predictions, which are obtained and plotted using Minitab in Figure 5.24 together with the 95% prediction interval.

Table 5.8 shows the output from JMP for fitting an AR(2) model to the weekly loan application data. In addition to the sample ACF and PACF, JMP provides the model fitting information including the estimates of the model parameters, the forecasts for 10 periods into the future and the associated prediction intervals, and the residual autocorrelation and PACF. The AR(2) model is an excellent fit to the data.

5.9 SEASONAL PROCESSES

Time series data may sometimes exhibit strong periodic patterns. This is often referred to as the time series having a seasonal behavior. This mostly occurs when data is taken in specific intervals—monthly, weekly, and so on. One way to represent such data is through an additive model where the process is assumed to be composed of two parts,

$$y_t = S_t + N_t \tag{5.102}$$

where S_t is the deterministic component with periodicity s and N_t is the stochastic component that may be modeled as an ARMA process. In that, y_t can be seen as a process with predictable periodic behavior with some noise sprinkled on top of it. Since the S_t is deterministic and has periodicity s, we have $S_t = S_{t+s}$ or

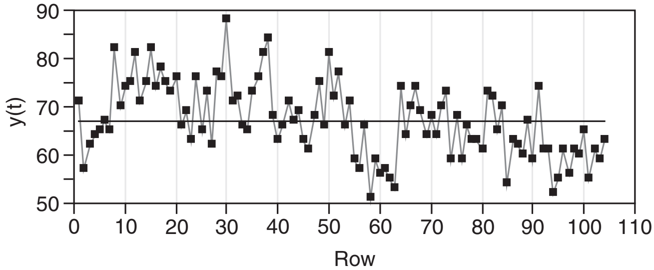
$$S_t - S_{t-s} = (1 - B^s)S_t = 0 \tag{5.103}$$

Applying the $(1 - B^s)$ operator to Eq. (5.102), we have

$$\begin{aligned} \underbrace{(1 - B^s)y_t}_{\equiv w_t} &= \underbrace{(1 - B^s)S_t}_{=0} + (1 - B^s)N_t \\ w_t &= (1 - B^s)N_t \end{aligned} \tag{5.104}$$

TABLE 5.8 JMP AR(2) Output for the Loan Application Data

Time series y(t)



Mean	67.067308
Std	7.663932
N	104
Zero Mean ADF	-0.695158
Single Mean ADF	-6.087814
Trend ADF	-7.396174

TABLE 5.8 (*Continued*)

Time series basic diagnostics

Lag	AutoCorr plot autocorr	Ljung-Box Q	p-Value
0	1.0000	.	.
1	0.4617	22.8186	<.0001
2	0.5314	53.3428	<.0001
3	0.2915	62.6167	<.0001
4	0.2682	70.5487	<.0001
5	0.2297	76.4252	<.0001
6	0.1918	80.5647	<.0001
7	0.2484	87.5762	<.0001
8	0.1162	89.1255	<.0001
9	0.1701	92.4847	<.0001
10	0.0565	92.8587	<.0001
11	0.0716	93.4667	<.0001
12	0.1169	95.1040	<.0001
13	0.1151	96.7080	<.0001
14	0.2411	103.829	<.0001
15	0.1137	105.430	<.0001
16	0.2540	113.515	<.0001
17	0.1279	115.587	<.0001
18	0.2392	122.922	<.0001
19	0.1138	124.603	<.0001
20	0.1657	128.206	<.0001
21	0.0745	128.944	<.0001
22	0.1320	131.286	<.0001
23	0.0708	131.968	<.0001
24	0.0338	132.125	<.0001
25	0.0057	132.130	<.0001

Lag	Partial plot partial		
Lag	**AutoCorr plot autocorr**		
0	1.0000	**Ljung-Box Q**	**p-Value**
1	0.4617		
2	0.4045		
3	–0.0629		
4	–0.0220		
5	0.0976		
6	0.0252		
7	0.1155		
8	–0.1017		
9	0.0145		
10	–0.0330		
11	–0.0250		
12	0.1349		
13	0.0488		
14	0.1489		
15	–0.0842		
16	0.1036		
17	0.0105		
18	0.0830		
19	–0.0938		
20	0.0052		
21	–0.0927		
22	0.1149		
23	–0.0645		
24	–0.0473		
25	–0.0742		

(*continued*)

TABLE 5.8 (*Continued*)

Model Comparison

Model	DF	Variance	AIC	SBC	RSquare	-2LogLH
AR(2)	101	39.458251	680.92398	688.85715	0.343	674.92398

Model: AR(2)
Model Summary

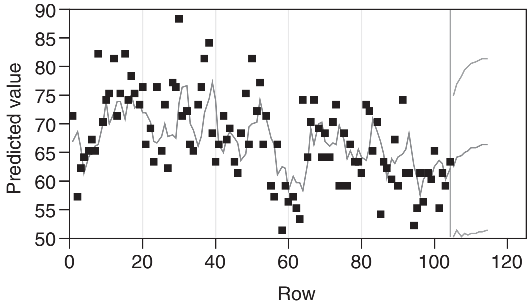
DF	101
Sum of Squared Errors	3985.28336
Variance Estimate	39.4582511
Standard Deviation	6.2815803
Akaike's 'A' Information Criterion	680.923978
Schwarz's Bayesian Criterion	688.857151
RSquare	0.34278547
RSquare Adj	0.32977132
MAPE	7.37857799
MAE	4.91939717
-2LogLikelihood	674.923978

Stable	Yes
Invertible	Yes

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	0.265885	0.089022	2.99	0.0035	21.469383
AR2	2	0.412978	0.090108	4.58	<.0001	
Intercept	0	66.854262	1.833390	36.46	<.0001	

Forecast



Residuals

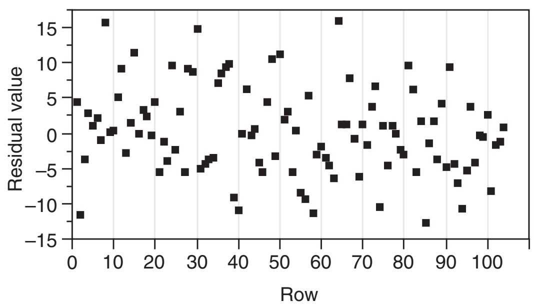


TABLE 5.8 (*Continued*)

Lag	AutoCorr	plot autocorr	Ljung-Box Q	p-Value
0	1.0000		.	.
1	0.0320		0.1094	0.7408
2	0.0287		0.1986	0.9055
3	–0.0710		0.7489	0.8617
4	–0.0614		1.1647	0.8839
5	–0.0131		1.1839	0.9464
6	0.0047		1.1864	0.9776
7	0.1465		3.6263	0.8217
8	–0.0309		3.7358	0.8801
9	0.0765		4.4158	0.8820
10	–0.0938		5.4479	0.8593
11	–0.0698		6.0251	0.8717
12	0.0019		6.0255	0.9148
13	0.0223		6.0859	0.9430
14	0.1604		9.2379	0.8155
15	–0.0543		9.6028	0.8440
16	0.1181		11.3501	0.7874
17	–0.0157		11.3812	0.8361
18	0.1299		13.5454	0.7582
19	–0.0059		13.5499	0.8093
20	0.0501		13.8788	0.8366
21	–0.0413		14.1056	0.8650
22	0.0937		15.2870	0.8496
23	0.0409		15.5146	0.8752
24	–0.0035		15.5163	0.9047
25	–0.0335		15.6731	0.9242

Lag	Partial	plot partial
0	1.0000	
1	0.0320	
2	0.0277	
3	–0.0729	
4	–0.0580	
5	–0.0053	
6	0.0038	
7	0.1399	
8	–0.0454	
9	0.0715	
10	–0.0803	

Lag	AutoCorr	plot autocorr	Ljung-Box Q	p-Value
11	–0.0586			
12	0.0201			
13	0.0211			
14	0.1306			
15	–0.0669			
16	0.1024			
17	0.0256			
18	0.1477			
19	–0.0027			
20	0.0569			
21	–0.0823			
22	0.1467			
23	–0.0124			
24	0.0448			
25	–0.0869			

The process w_t can be seen as **seasonally stationary**. Since an ARMA process can be used to model N_t, in general, we have

$$\Phi(B)w_t = (1 - B^s)\Theta(B)\varepsilon_t \tag{5.105}$$

where ε_t is white noise.

We can also consider S_t as a stochastic process. We will further assume that after seasonal differencing, $(1 - B^s)$, $(1 - B^s)\,y_t = w_t$ becomes stationary. This, however, may not eliminate all seasonal features in the process. That is, the seasonally differenced data may still show strong autocorrelation at lags s, $2s$, … . So the seasonal ARMA model is

$$\left(1 - \phi_1^* B^s - \phi_2^* B^{2s} - \cdots - \phi_P^* B^{Ps}\right) w_t = \left(1 - \theta_1^* B^s - \theta_2^* B^{2s} - \cdots - \theta_Q^* B^{Qs}\right) \varepsilon_t \tag{5.106}$$

This representation, however, only takes into account the autocorrelation at seasonal lags s, $2s$, … . Hence a more general seasonal ARIMA model of orders $(p, d, q) \times (P, D, Q)$ with period s is

$$\Phi^*(B^s)\Phi(B)(1 - B)^d(1 - B^s)^D y_t = \delta + \Theta^*(B^s)\Theta(B)\varepsilon_t \tag{5.107}$$

In practice, although it is case specific, it is not expected to have P, D, and Q greater than 1. The results for regular ARIMA processes that we discussed in previous sections apply to the seasonal models given in Eq. (5.107).

As in the nonseasonal ARIMA models, the forecasts for the seasonal ARIMA models can be obtained from the difference equations as illustrated for example in Eq. (5.101) for a nonseasonal ARIMA(1,1,1) process. Similarly the weights in the random shock form given in Eq. (5.96) can be estimated as in Eq. (5.98) to obtain the estimate for the variance of the forecast errors as well as the prediction intervals given in Eqs. (5.91) and (5.92), respectively.

Example 5.7 The ARIMA (0, 1, 1) × (0, 1, 1) model with $s = 12$ is

$$\underbrace{(1 - B)(1 - B^{12})y_t}_{w_t} = \left(1 - \theta_1 B - \theta_1^* B^{12} + \theta_1 \theta_1^* B^{13}\right)\varepsilon_t$$

For this process, the autocovariances are calculated as

$$\begin{aligned}
\gamma(0) &= \mathrm{Var}(w_t) = \sigma^2\left(1+\theta_1^2+\theta_1^{*2}+\left(-\theta_1\theta_1^*\right)^2\right)\\
&= \sigma^2\left(1+\theta_1^2\right)\left(1+\theta_1^{*2}\right)\\
\gamma(1) &= \mathrm{Cov}(w_t, w_{t-1}) = \sigma^2\left(-\theta_1+\theta_1^*\left(-\theta_1\theta_1^*\right)\right)\\
&= -\theta_1\sigma^2\left(1+\theta_1^*\right)\\
\gamma(2) &= \gamma(3) = \cdots = \gamma(10) = 0\\
\gamma(11) &= \sigma^2\theta_1\theta_1^*\\
\gamma(12) &= -\sigma^2\theta_1^*\left(1+\theta_1^2\right)\\
\gamma(13) &= \sigma^2\theta_1\theta_1^*\\
\gamma(j) &= 0, \quad j > 13
\end{aligned}$$

Example 5.8 Consider the US clothing sales data in Table 4.9. The data obviously exhibit some seasonality and upward linear trend. The sample ACF and PACF plots given in Figure 5.25 indicate a monthly seasonality, $s = 12$, as ACF values at lags 12, 24, 36 are significant and slowly decreasing, and there is a significant PACF value at lag 12 that is close to 1. Moreover, the slowly decreasing ACF in general, also indicates a nonstationarity that can be remedied by taking the first difference. Hence we would now consider $w_t = (1-B)(1-B^{12})y_t$.

Figure 5.26 shows that first difference together with seasonal differencing—that is, $w_t = (1-B)(1-B^{12})y_t$—helps in terms of stationarity and eliminating the seasonality, which is also confirmed by sample ACF and PACF plots given in Figure 5.27. Moreover, the sample ACF with a significant value at lag 1 and the sample PACF with exponentially

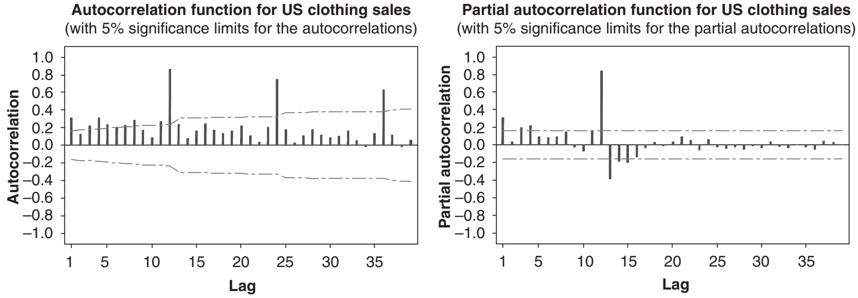


FIGURE 5.25 Sample ACF and PACF plots of the US clothing sales data.

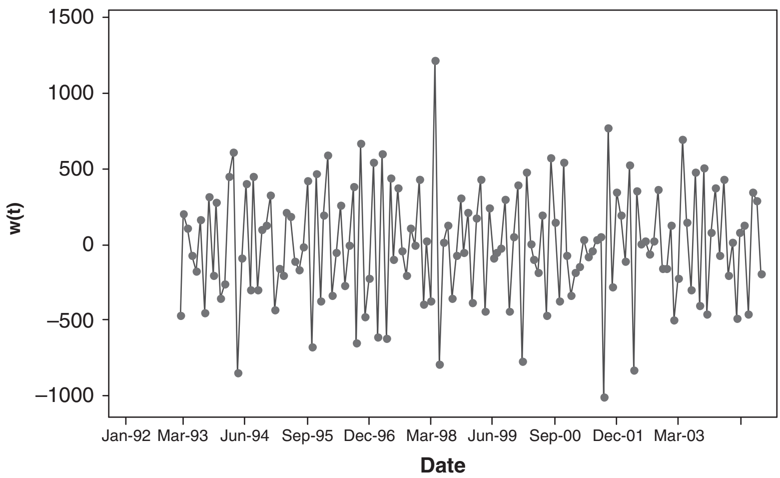


FIGURE 5.26 Time series plot of $w_t = (1 - B)(1 - B^{12})y_t$ for the US clothing sales data.

decaying values at the first 8 lags suggest that a nonseasonal MA(1) model should be used.

The interpretation of the remaining seasonality is a bit more difficult. For that we should focus on the sample ACF and PACF values at lags 12, 24, 36, and so on. The sample ACF at lag 12 seems to be significant and the sample PACF at lags 12, 24, 36 (albeit not significant) seems to be alternating in sign. That suggests that a seasonal MA(1) model can be used as well. Hence an ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model is used to model the data, y_t. The output from Minitab is given in Table 5.9. Both MA(1) and seasonal MA(1) coefficient estimates are significant. As we can see from the sample ACF and PACF plots in Figure 5.28, while there are still some

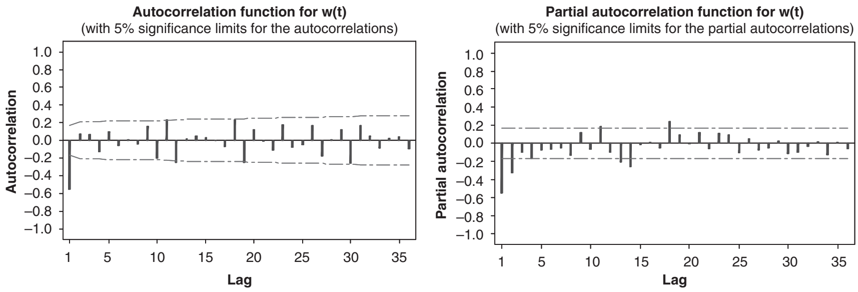


FIGURE 5.27 Sample ACF and PACF plots of $w_t = (1 - B)(1 - B^{12})y_t$.

TABLE 5.9 Minitab Output for the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ Model for the US Clothing Sales Data

```
Final Estimates of Parameters

Type        Coef   SE Coef       T      P
MA   1    0.7626    0.0542   14.06  0.000
SMA  12   0.5080    0.0771    6.59  0.000


Differencing: 1 regular, 1 seasonal of order 12
Number of observations:  Original series 155, after
differencing 142
Residuals:    SS =  10033560 (backforecasts excluded)
              MS =  71668   DF = 140


Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag             12     24     36     48
Chi-Square    15.8   37.7   68.9   92.6
DF              10     22     34     46
P-Value      0.107  0.020  0.000  0.000
```

small significant values, as indicated by the modified Box pierce statistic most of the autocorrelation is now modeled out.

The residual plots in Figure 5.29 provided by Minitab seem to be acceptable as well.

Finally, the time series plot of the actual and fitted values in Figure 5.30 suggests that the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model provides a reasonable fit to this highly seasonal and nonstationary time series data.

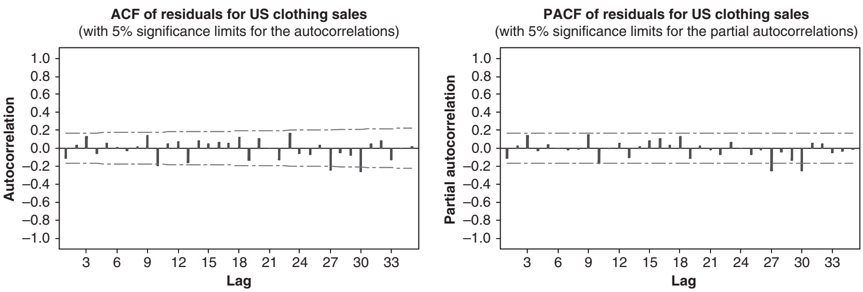


FIGURE 5.28 Sample ACF and PACF plots of residuals from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model.

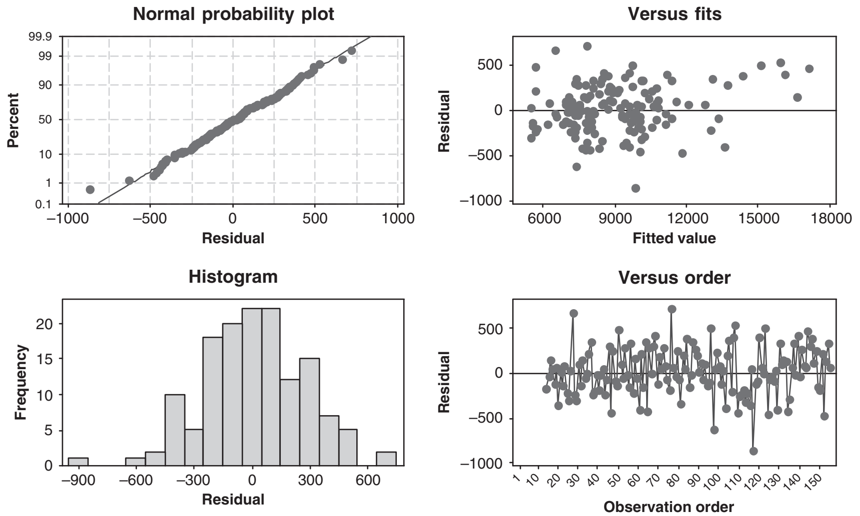


FIGURE 5.29 Residual plots from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model for the US clothing sales data.

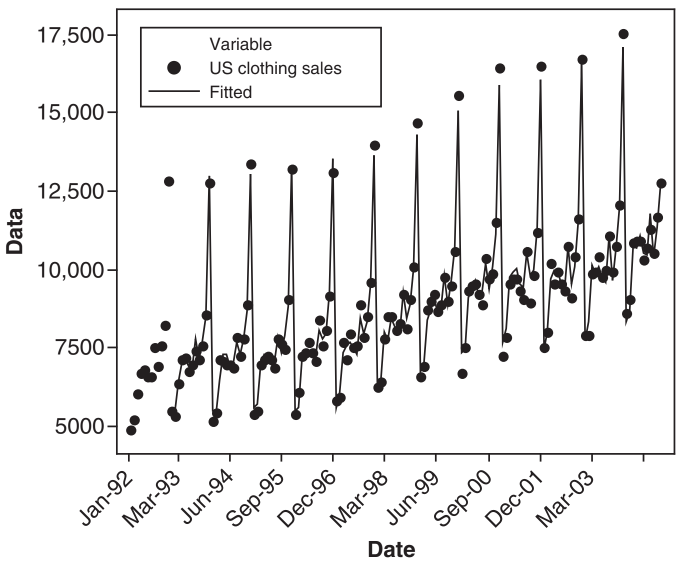


FIGURE 5.30 Time series plot of the actual data and fitted values from the ARIMA(0, 1, 1) × (0, 1, 1)$_{12}$ model for the US clothing sales data.

5.10 ARIMA MODELING OF BIOSURVEILLANCE DATA

In Section 4.8 we introduced the daily counts of respiratory and gastrointestinal complaints for more than 2-$^1/_2$ years at several hospitals in a large metropolitan area from Fricker (2013). Table 4.12 presents the 980 observations from one of these hospitals. Section 4.8 described modeling the respiratory count data with exponential smoothing. We now present an ARIMA modeling approach. Figure 5.31 presents the sample ACF, PACF, and the variogram from JMP for these data. Examination of the original time series plot in Figure 4.35 and the ACF and variogram indicate that the daily respiratory syndrome counts may be nonstationary and that the data should be differenced to obtain a stationary time series for ARIMA modeling.

The ACF for the differenced series ($d = 1$) shown in Figure 5.32 cuts off after lag 1 while the PACF appears to be a mixture of exponential decays. This suggests either an ARIMA(1, 1, 1) or ARIMA(2, 1, 1) model.

The Time Series Modeling platform in JMP allows a group of ARIMA models to be fit by specifying ranges for the AR, difference, and MA terms. Table 5.10 summarizes the fits obtained for a constant difference (d = 1), and both AR (p) and MA (q) parameters ranging from 0 to 2.

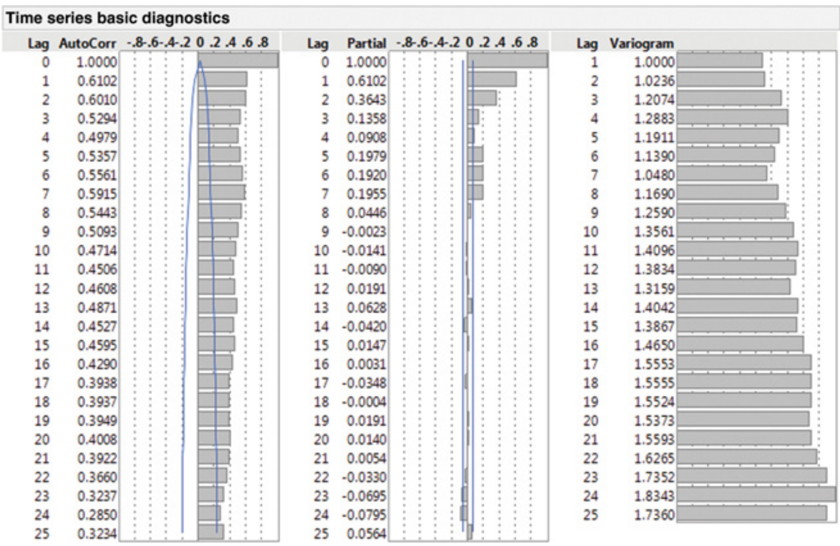


FIGURE 5.31 ACF, PACF, and variogram for daily respiratory syndrome counts.

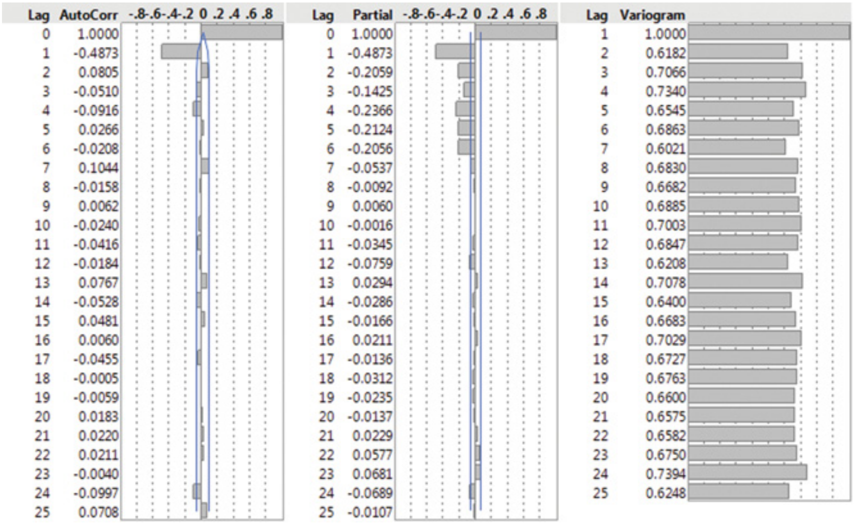


FIGURE 5.32 ACF, PACF, and variogram for the first difference of the daily respiratory syndrome counts.

TABLE 5.10 Summary of Models fit to the Respiratory Syndrome Count Data

Model	Variance	AIC	BIC	RSquare	MAPE	MAE
AR(1)	65.7	6885.5	6895.3	0.4	24.8	6.3
AR(2)	57.1	6748.2	6762.9	0.5	22.9	5.9
MA(1)	81.6	7096.9	7106.6	0.2	28.5	6.9
MA(2)	69.3	6937.6	6952.3	0.3	26.2	6.4
ARMA(1, 1)	52.2	6661.2	6675.9	0.5	21.6	5.6
ARMA(1, 2)	52.1	6661.2	6680.7	0.5	21.6	5.6
ARMA(2, 1)	52.1	6660.7	6680.3	0.5	21.6	5.6
ARMA(2, 2)	52.3	6664.3	6688.7	0.5	21.6	5.6
ARIMA(0, 0, 0)	104.7	7340.4	7345.3	0.0	33.2	8.0
ARIMA(0, 1, 0)*	81.6	7088.2	7093.1	0.2	26.2	7.0
ARIMA(0, 1, 1)*	52.7	6662.8	6672.6	0.5	21.4	5.7
ARIMA(0, 1, 2)	52.6	6662.1	6676.7	0.5	21.4	5.7
ARIMA(1, 1, 0)*	62.2	6824.4	6834.2	0.4	23.2	6.2
ARIMA(1, 1, 1)	52.6	6661.4	6676.1	0.5	21.4	5.7
ARIMA(1, 1, 2)	52.6	6661.9	6681.5	0.5	21.4	5.7
ARIMA(2, 1, 0)	59.6	6783.5	6798.1	0.4	22.7	6.1
ARIMA(2, 1, 1)	52.3	6657.1	6676.6	0.5	21.4	5.6
ARIMA(2, 1, 2)	52.3	6657.8	6682.2	0.5	21.3	5.6

*Indicates that objective function failed during parameter estimation.

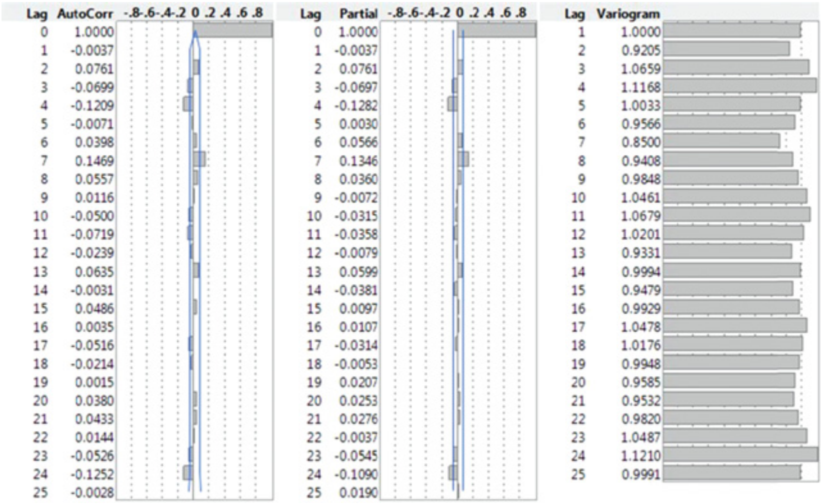


FIGURE 5.33 ACF, PACF, and variogram for the residuals of ARIMA(1, 1, 1) fit to daily respiratory syndrome counts.

In terms of the model summary statistics variance of the errors, AIC and mean absolute prediction error (MAPE) several models look potentially reasonable. For the ARIMA(1, 1, 1) we obtained the following results from JMP:

Parameter estimates

Term	Lag	Estimate	Std error	t Ratio	Prob>|t|	Constant estimate
AR1	1	0.07307009	0.0394408	1.85	0.0642	0.00069557
MA1	1	0.81584055	0.0223680	36.47	<.0001*	
Intercept	0	0.00075040	0.0036018	0.21	0.8350	

Figure 5.33 presents the ACF, PACF, and variogram of the residuals from this model. Other residual plots are in Figure 5.34.

For comparison purposes we also fit the ARIMA(2, 1, 1) model. The parameter estimates obtained from JMP are:

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	0.09953471	0.0402040	2.48	0.0135*	0.00097496
AR2	2	0.09408008	0.0375486	2.51	0.0124*	
MA1	1	0.84755625	0.0231814	36.56	<.0001*	
Intercept	0	0.00120905	0.0088678	0.14	0.8916	

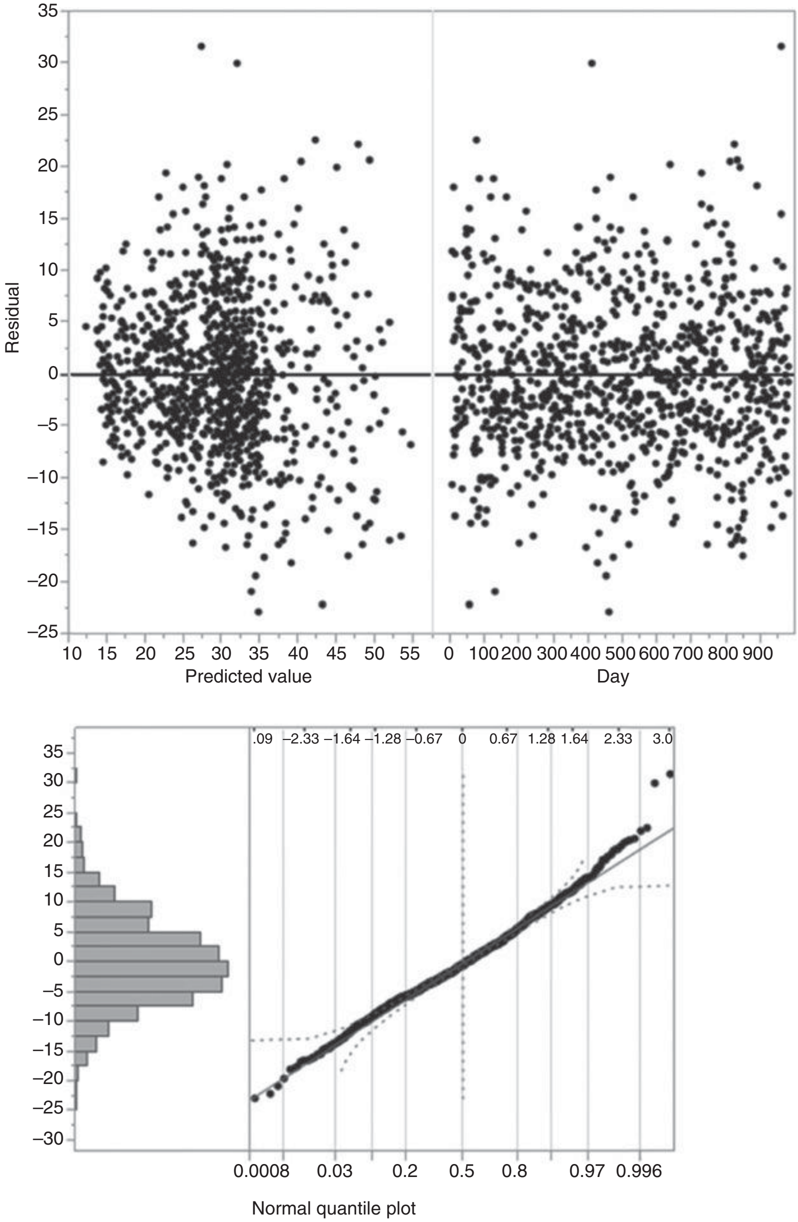


FIGURE 5.34 Plots of residuals from ARIMA(1, 1, 1) fit to daily respiratory syndrome counts.

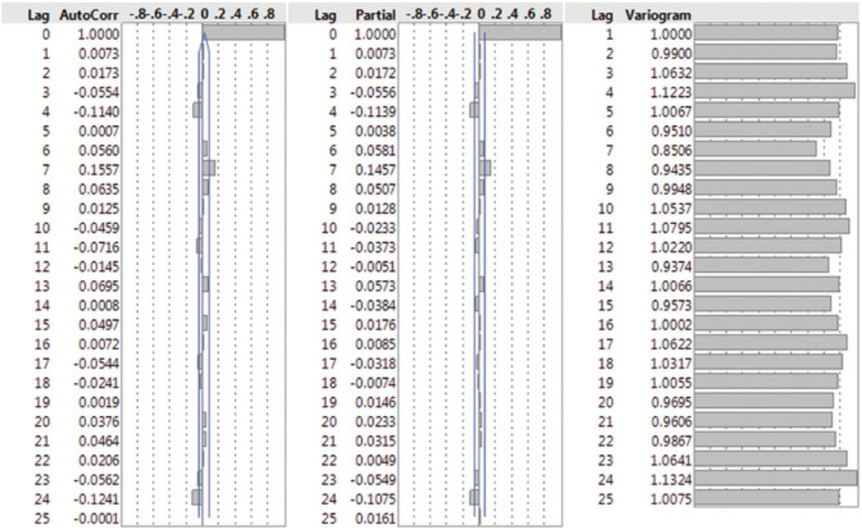


FIGURE 5.35 ACF, PACF, and variogram for residuals of ARIMA(2, 1, 1) fit to daily respiratory syndrome counts.

The lag 2 AR parameter is highly significant. Figure 5.35 presents the plots of the ACF, PACF, and variogram of the residuals from ARIMA(2, 1, 1). Other residual plots are shown in Figure 5.36. Based on the significant lag 2 AR parameter, this model is preferable to the ARIMA(1, 1, 1) model fit previously.

Considering the variation in counts by day of week that was observed previously, a seasonal ARIMA model with a seasonal period of 7 days may be appropriate. The resulting model has an error variance of 50.9, smaller than for the ARIMA(1, 1, 1) and ARIMA(2, 1, 1) models. The AIC is also smaller. Notice that all of the model parameters are highly significant. The residual ACF, PACF, and variogram shown in Figure 5.37 do not suggest any remaining structure. Other residual plots are in Figure 5.38.

Model	Variance	AIC	BIC	RSquare	MAPE	MAE
ARIMA(2, 1, 1)(0, 0, 1)$_7$	50.9	6631.9	6656.4	0.5	21.1	5.6

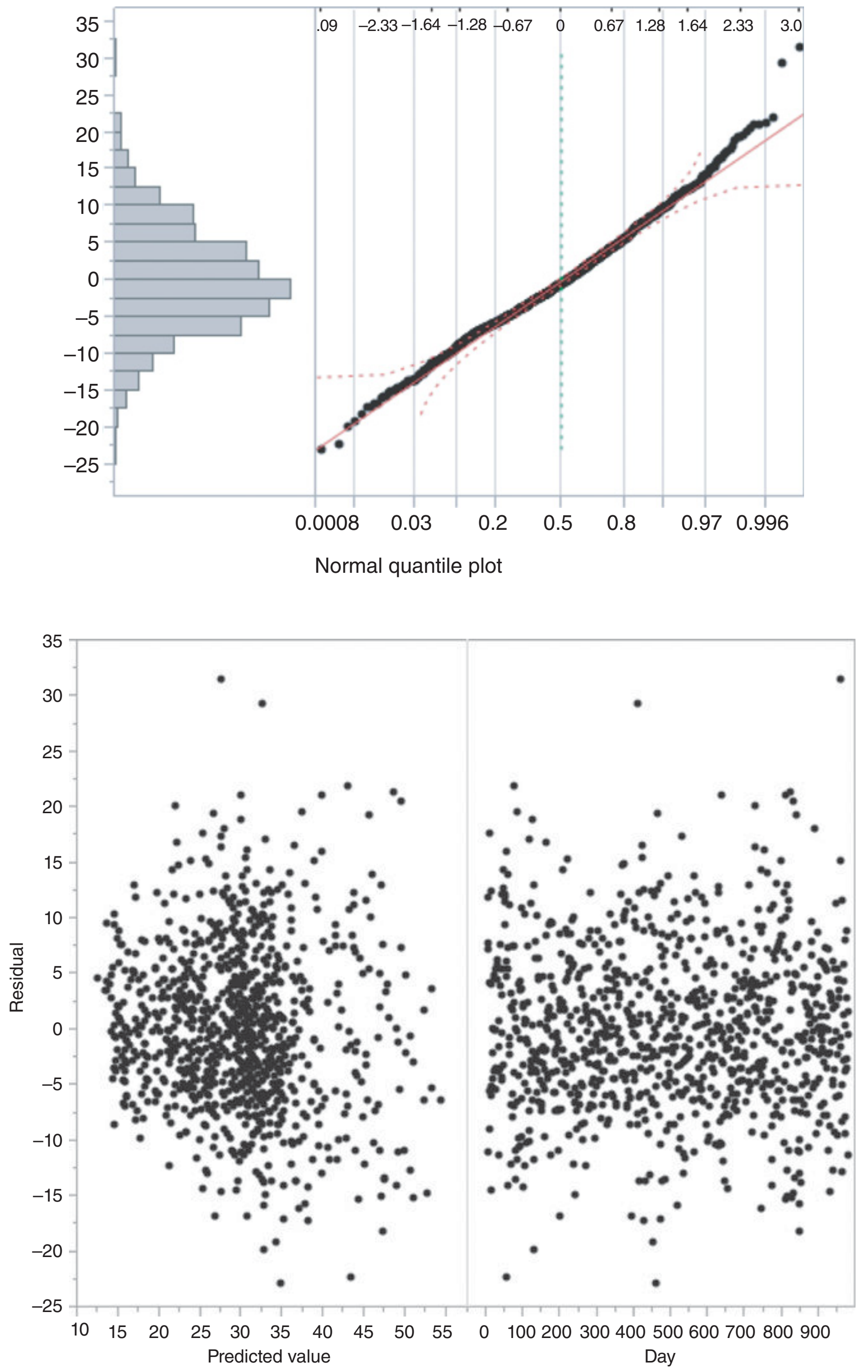
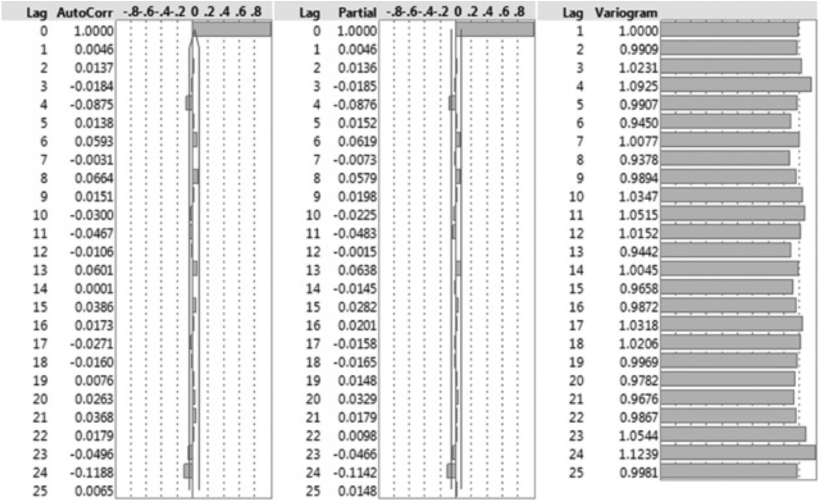


FIGURE 5.36 Plots of residuals from ARIMA(2, 1, 1) fit to daily respiratory syndrome counts.

Parameter estimates							
Term	Factor	Lag	Estimate	Std error	t Ratio	Prob>|t|	Constant estimate
AR1,1	1	1	0.1090685	0.0395388	2.76	0.0059*	
AR1,2	1	2	0.1186083	0.0376471	3.15	0.0017*	0.00083453
MA1,1	1	1	0.8730535	0.0225127	38.78	<.0001*	
MA2,7	2	7	-0.1744415	0.0328363	-5.31	<.0001*	
Intercept	1	0	0.0010805	0.0051887	0.21	0.8351	



Lag	AutoCorr	Lag	Partial	Lag	Variogram
0	1.0000	0	1.0000	1	1.0000
1	0.0046	1	0.0046	2	0.9909
2	0.0137	2	0.0136	3	1.0231
3	-0.0184	3	-0.0185	4	1.0925
4	-0.0875	4	-0.0876	5	0.9907
5	0.0138	5	0.0152	6	0.9450
6	0.0593	6	0.0619	7	1.0077
7	-0.0031	7	-0.0073	8	0.9378
8	0.0664	8	0.0579	9	0.9894
9	0.0151	9	0.0198	10	1.0347
10	-0.0300	10	-0.0225	11	1.0515
11	-0.0467	11	-0.0483	12	1.0152
12	-0.0106	12	-0.0015	13	0.9442
13	0.0601	13	0.0638	14	1.0045
14	0.0001	14	-0.0145	15	0.9658
15	0.0386	15	0.0282	16	0.9872
16	0.0173	16	0.0201	17	1.0318
17	-0.0271	17	-0.0158	18	1.0206
18	-0.0160	18	-0.0165	19	0.9969
19	0.0076	19	0.0148	20	0.9782
20	0.0263	20	0.0329	21	0.9676
21	0.0368	21	0.0179	22	0.9867
22	0.0179	22	0.0098	23	1.0544
23	-0.0496	23	-0.0466	24	1.1239
24	-0.1188	24	-0.1142	25	0.9981
25	0.0065	25	0.0148		

FIGURE 5.37 ACF, PACF, and variogram of residuals from ARIMA(2, 1, 1) × (0, 0, 1)$_7$ fit to daily respiratory syndrome counts.

5.11 FINAL COMMENTS

ARIMA models (a.k.a. Box–Jenkins models) present a very powerful and flexible class of models for time series analysis and forecasting. Over the years, they have been very successfully applied to many problems in research and practice. However, there might be certain situations where they may fall short on providing the "right" answers. For example, in ARIMA models, forecasting future observations primarily relies on the past data and implicitly assumes that the conditions at which the data is collected will remain the same in the future as well. In many situations this assumption may (and most likely will) not be appropriate. For those cases, the transfer function–noise models, where a set of input variables that may have an effect on the time series are added to the model, provide suitable options. We shall discuss these models in the next chapter. For an excellent discussion of this matter and of time series analysis and forecasting in general, see Jenkins (1979).

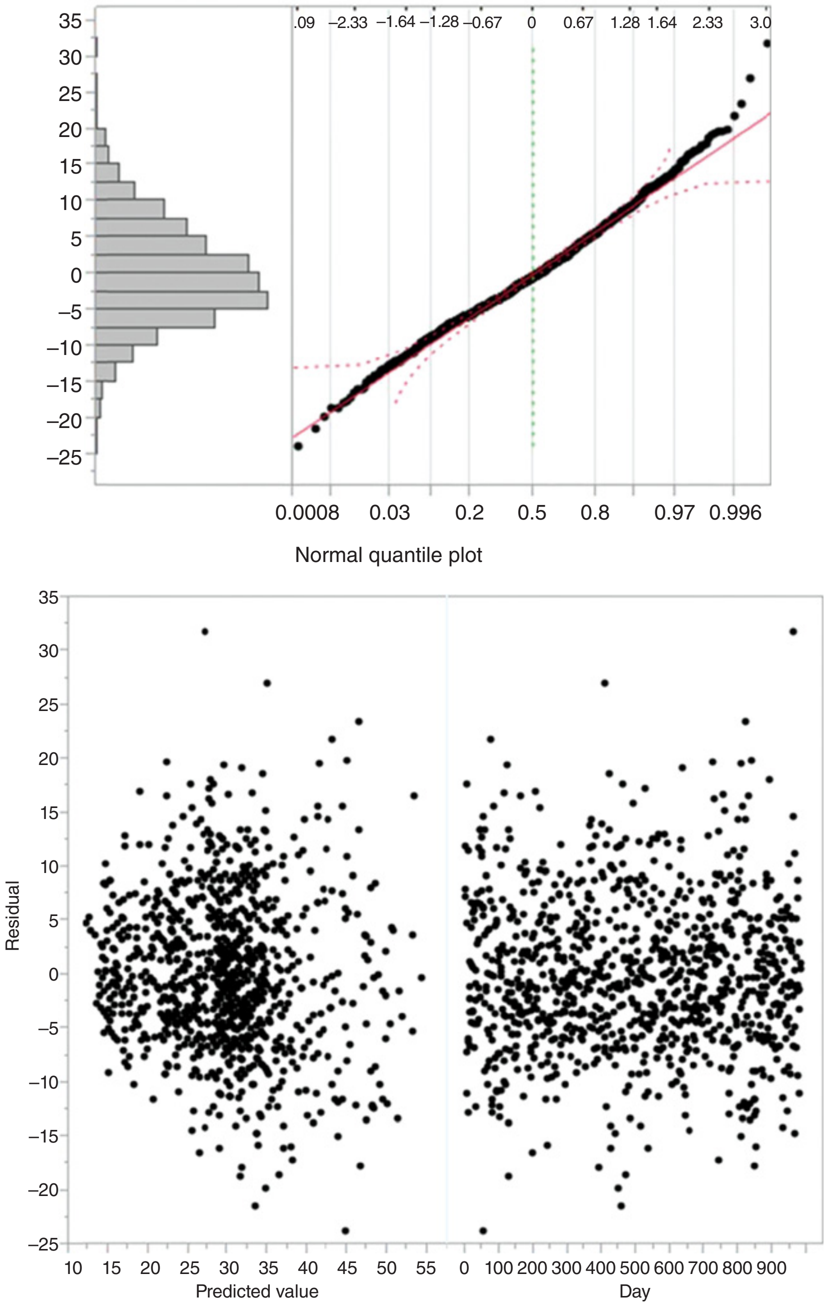
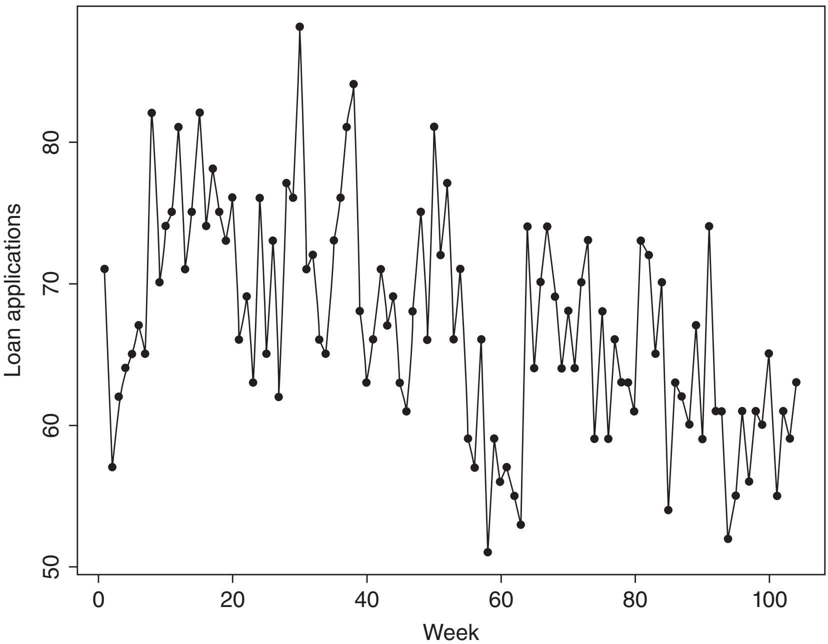


FIGURE 5.38 Plots of residuals from ARIMA(2, 1, 1) × (0, 0, 1)$_7$ fit to daily respiratory syndrome counts.

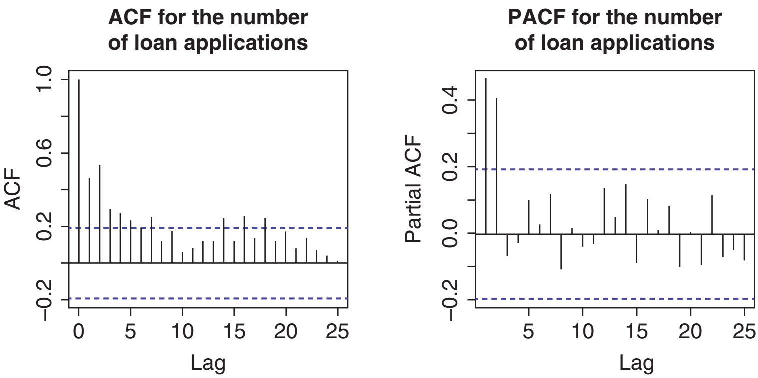
5.12 R COMMANDS FOR CHAPTER 5

Example 5.1 The loan applications data are in the second column of the array called loan.data in which the first column is the number of weeks. We first plot the data as well as the ACF and PACF.

```
plot(loan.data[,2],type="o",pch=16,cex=.5,xlab='Week',ylab='Loan
Applications')
```



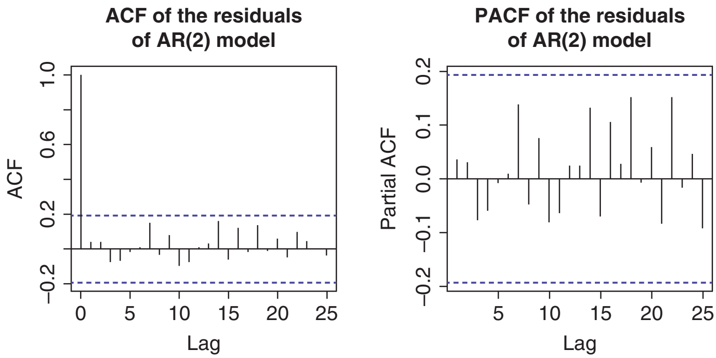

```
par(mfrow=c(1,2),oma=c(0,0,0,0))

acf(loan.data[,2],lag.max=25,type="correlation",main="ACF for the
Number \nof Loan Applications")

acf(loan.data[,2], lag.max=25,type="partial",main="PACF for the
Number \nof Loan Applications")
```



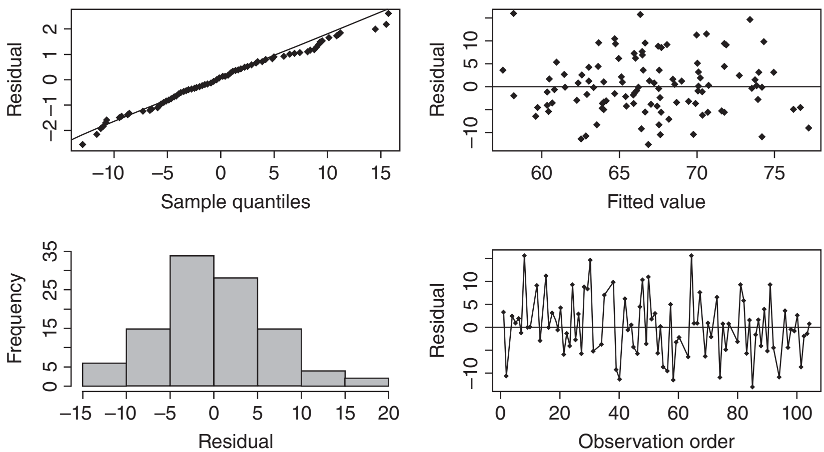
Fit an ARIMA(2,0,0) model to the data using arima function in the stats package.

```
loan.fit.ar2<-arima(loan.data[,2],order=c(2, 0, 0))
loan.fit.ar2

       Call:
       arima(x = loan.data[, 2], order = c(2, 0, 0))

       Coefficients:
                ar1     ar2  intercept
             0.2659  0.4130    66.8538
       s.e.  0.0890  0.0901     1.8334

       sigma^2 estimated as 38.32:  log likelihood = -337.46,
       aic = 682.92


res.loan.ar2<-as.vector(residuals(loan.fit.ar2))
#to obtain the fitted values we use the function fitted() from
#the forecast package
library(forecast)
fit.loan.ar2<-as.vector(fitted(loan.fit.ar2))

Box.test(res.loan.ar2,lag=48,fitdf=3,type="Ljung")

                Box-Ljung test

       data:  res.loan.ar2
       X-squared = 31.8924, df = 45, p-value = 0.9295


#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
```

```
acf(res.loan.ar2,lag.max=25,type="correlation",main="ACF of the
Residuals \nof AR(2) Model")

acf(res.loan.ar2, lag.max=25,type="partial",main="PACF of the
Residuals \nof AR(2) Model")
```

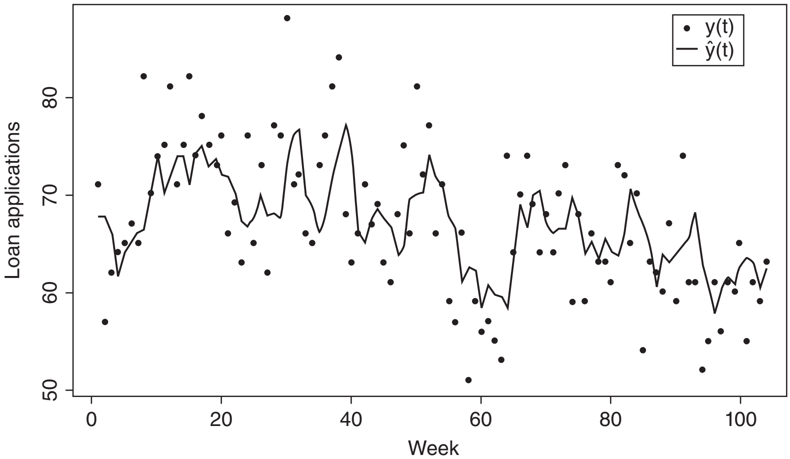



```
#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.loan.ar2,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.loan.ar2,datax=TRUE)
plot(fit.loan.ar2,res.loan.ar2,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(res.loan.ar2,col="gray",xlab='Residual',main='')
plot(res.loan.ar2,type="l",xlab='Observation Order',
ylab='Residual')
points(res.loan.ar2,pch=16,cex=.5)
abline(h=0)
```



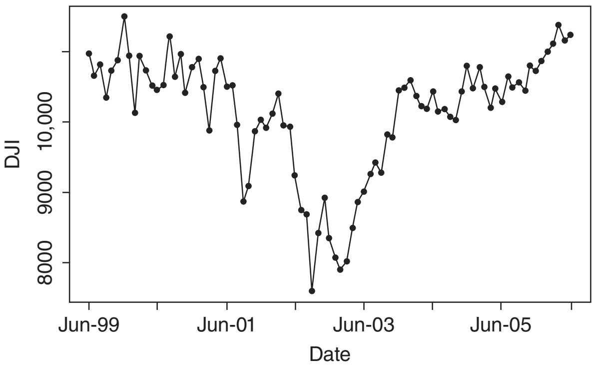
Plot fitted values

```
plot(loan.data[,2],type="p",pch=16,cex=.5,xlab='Week',ylab='Loan
Applications')
lines(fit.loan.ar2)
legend(95,88,c("y(t)","yhat(t)"), pch=c(16, NA),lwd=c(NA,.5),
cex=.55)
```

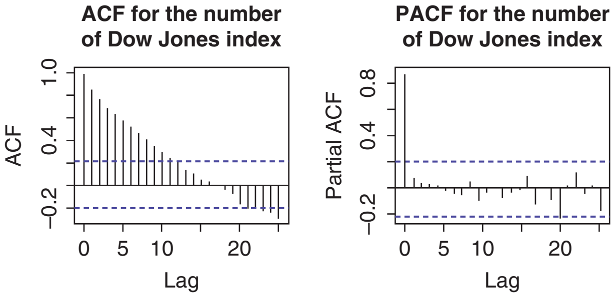


Example 5.2 The Dow Jones index data are in the second column of the array called dji.data in which the first column is the month of the year. We first plot the data as well as the ACF and PACF.

```
plot(dji.data[,2],type="o",pch=16,cex=.5,xlab='Date',ylab='DJI',
xaxt='n')
axis(1, seq(1,85,12), dji.data[seq(1,85,12),1])
```



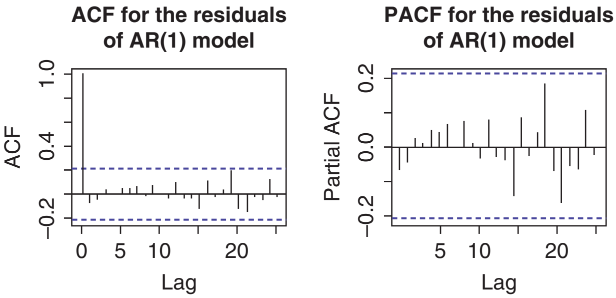
```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(dji.data[,2],lag.max=25,type="correlation",main="ACF for the
Number \nof Dow Jones Index")

acf(dji.data[,2], lag.max=25,type="partial",main="PACF for the
Number \nof Dow Jones Index ")
```



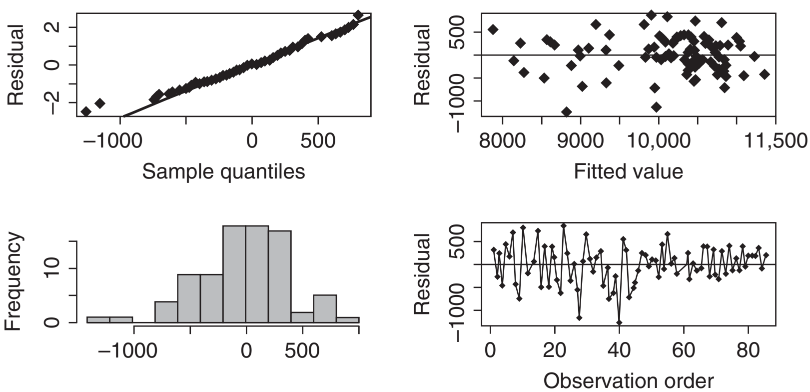
We first fit an ARIMA(1,0,0) model to the data using arima function in the stats package.

```
dji.fit.ar1<-arima(dji.data[,2],order=c(1, 0, 0))
dji.fit.ar1
       Call:
       arima(x = dji.data[, 2], order = c(1, 0, 0))

       Coefficients:
                ar1   intercept
             0.8934  10291.2984
       s.e.  0.0473    373.8723

       sigma^2 estimated as 156691:  log likelihood = -629.8,
       aic = 1265.59


res.dji.ar1<-as.vector(residuals(dji.fit.ar1))
#to obtain the fitted values we use the function fitted() from
#the forecast package
library(forecast)
fit.dji.ar1<-as.vector(fitted(dji.fit.ar1))

Box.test(res.dji.ar1,lag=48,fitdf=3,type="Ljung")

        Box-Ljung test

        data:  res.dji.ar1
        X-squared = 29.9747, df = 45, p-value = 0.9584
```

```
#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.dji.ar1,lag.max=25,type="correlation",main="ACF of the
Residuals \nof AR(1) Model")

acf(res.dji.ar1, lag.max=25,type="partial",main="PACF of the
Residuals \nof AR(1) Model")
```

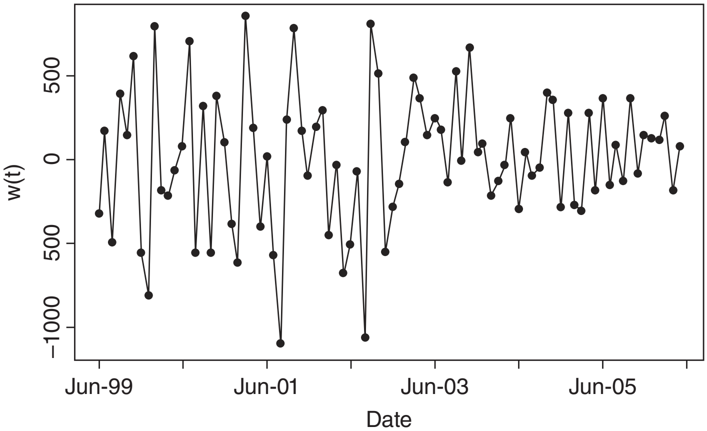



```
#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.dji.ar1,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.dji.ar1,datax=TRUE)
plot(fit.dji.ar1,res.dji.ar1,pch=16, xlab='Fitted Value',
ylab='Residual')
abline(h=0)
hist(res.dji.ar1,col="gray",xlab='Residual',main='')
plot(res.dji.ar1,type="l",xlab='Observation Order',
ylab='Residual')
points(res.dji.ar1,pch=16,cex=.5)
abline(h=0)
```

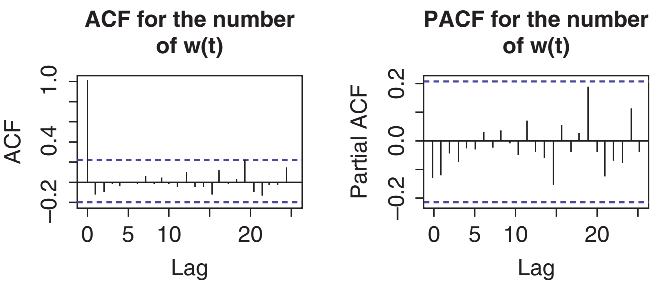


We now consider the first difference of the Dow Jones index.

```
wt.dji<-diff(dji.data[,2])
plot(wt.dji,type="o",pch=16,cex=.5,xlab='Date',ylab='w(t)',
xaxt='n')
axis(1, seq(1,85,12), dji.data[seq(1,85,12),1])
```




```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(wt.dji,lag.max=25,type="correlation",main="ACF for the
Number \nof w(t)")

acf(wt.dji, lag.max=25,type="partial",main="PACF for the
Number \nof w(t)")
```



Example 5.6 The loan applications data are in the second column of the array called loan.data in which the first column is the number of weeks. We use the AR(2) model to make the forecasts.

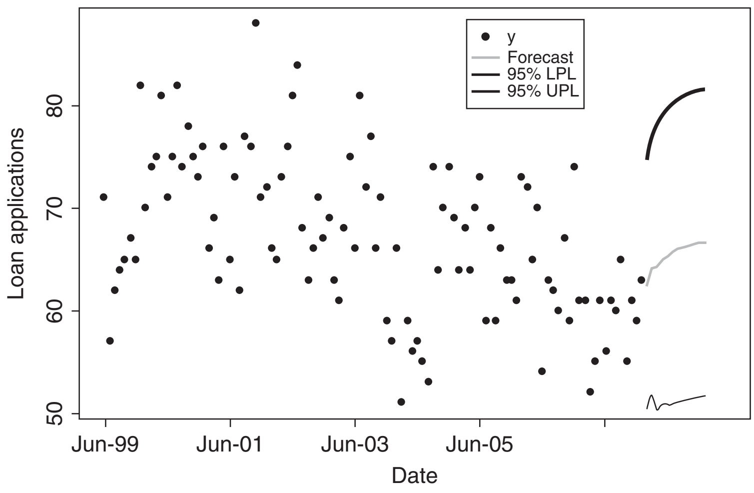
```
loan.fit.ar2<-arima(loan.data[,2],order=c(2, 0, 0))
#to obtain the 1- to 12-step ahead forecasts, we use the
#function forecast() from the forecast package
library(forecast)
loan.ar2.forecast<-as.array(forecast(loan.fit.ar2,h=12))
loan.ar2.forecast
    Point  Forecast     Lo 80     Hi 80     Lo 95     Hi 95
105        62.58571  54.65250  70.51892  50.45291  74.71851
106        64.12744  55.91858  72.33629  51.57307  76.68180
107        64.36628  55.30492  73.42764  50.50812  78.22444
108        65.06647  55.80983  74.32312  50.90965  79.22330
109        65.35129  55.86218  74.84039  50.83895  79.86362
110        65.71617  56.13346  75.29889  51.06068  80.37167
111        65.93081  56.27109  75.59054  51.15754  80.70409
112        66.13857  56.43926  75.83789  51.30475  80.97240
113        66.28246  56.55529  76.00962  51.40605  81.15887
114        66.40651  56.66341  76.14961  51.50572  81.30730
115        66.49892  56.74534  76.25249  51.58211  81.41572
116        66.57472  56.81486  76.33458  51.64830  81.50114
```

Note that forecast function provides a list with forecasts as well as 80% and 95% prediction limits. To see the elements of the list, we can do

```
ls(loan.ar2.forecast)
       [1] "fitted"     "level"  "lower"  "mean"  "method"  "model"
       [7] "residuals"  "upper"  "x"      "xname"
```

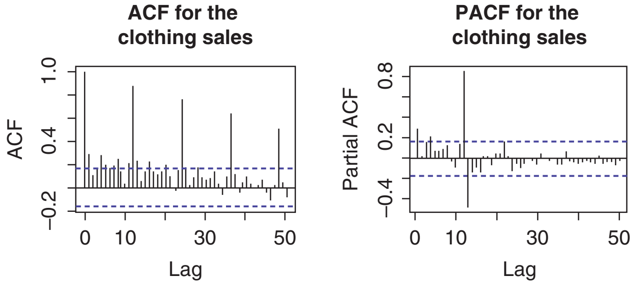
In this list, “mean” stands for the forecasts while “lower” and “upper” provide the 80 and 95% lower and upper prediction limits, respectively. To plot the forecasts and the prediction limits, we have

```
plot(loan.data[,2],type="p",pch=16,cex=.5,xlab='Date',ylab='Loan
Applications',xaxt='n',xlim=c(1,120))
axis(1, seq(1,120,24), dji.data[seq(1,120,24),1])
lines(105:116,loan.ar2.forecast$mean,col="grey40")
lines(105:116,loan.ar2.forecast$lower[,2])
lines(105:116,loan.ar2.forecast$upper[,2])
legend(72,88,c("y","Forecast","95% LPL","95% UPL"), pch=c(16, NA,
NA,NA),lwd=c(NA,.5,.5,.5),cex=.55,col=c("black","grey40","black",
"black"))
```



Example 5.8 The clothing sales data are in the second column of the array called closales.data in which the first column is the month of the year. We first plot the data and its ACF and PACF.

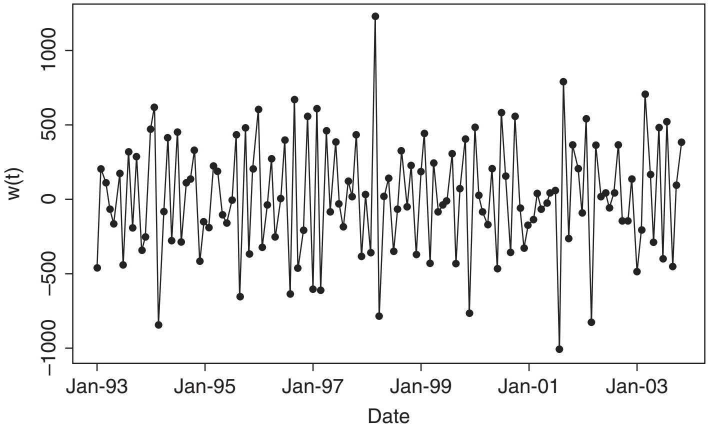
```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(closales.data[,2],lag.max=50,type="correlation",main="ACF for
the \n Clothing Sales")

acf(closales.data[,2], lag.max=50,type="partial",main="PACF for
the \n Clothing Sales")
```



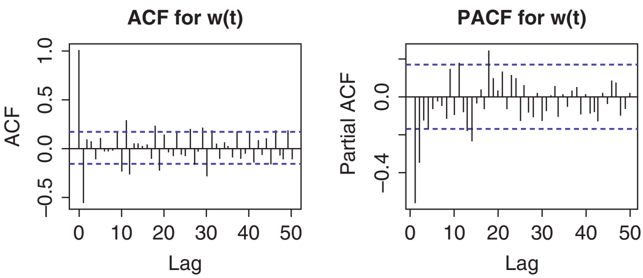
We now take the seasonal and non-seasonal difference of the data.

```
wt.closales<-diff(diff(closales.data[,2],lag=1),lag=12)
#Note that the same result would have been obtained with the
#following command when the order of differencing is reversed
#wt.closales<-diff(diff(closales.data[,2],lag=12),lag=1)
```

```
plot(wt.closales,type="o",pch=16,cex=.5,xlab='Date',ylab='w(t)',
xaxt='n')
axis(1, seq(1,144,24), closales.data[seq(13,144,24),1])
```



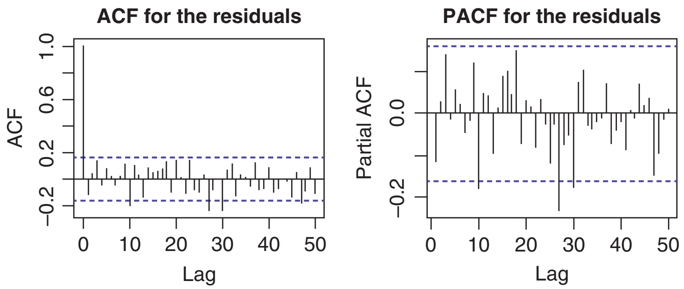

```
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(wt.closales,lag.max=50,type="correlation",main="ACF for w(t)")
acf(wt.closales, lag.max=50,type="partial",main="PACF for w(t)")
```



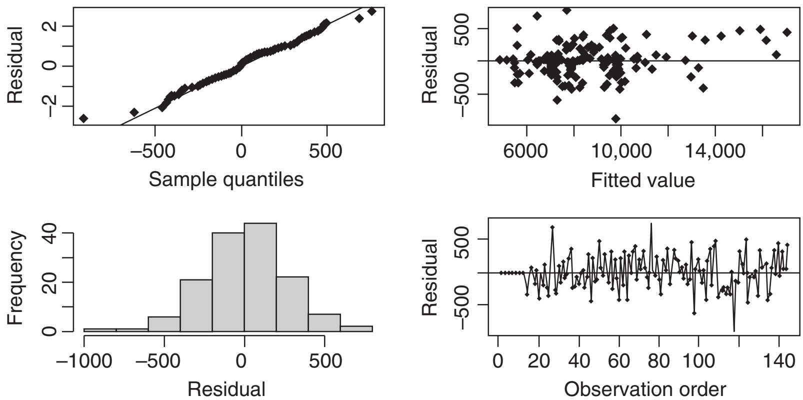
We now fit a seasonal ARIMA(0,1,1) × $(0,1,1)_{12}$ model to the data. We then plot the residuals plots including ACF and PACF of the residuals. In the end we plot the true and fitted values.

```
closales.fit.sar<-arima(closales.data[,2],order=c(0,1,1),
seasonal=list(order = c(0,1,1),period=12),)

res.closales.sar<-as.vector(residuals(closales.fit.sar))
#to obtain the fitted values we use the function fitted() from
the forecast package
library(forecast)
fit.closales.sar<-as.vector(fitted(closales.fit.sar))
```

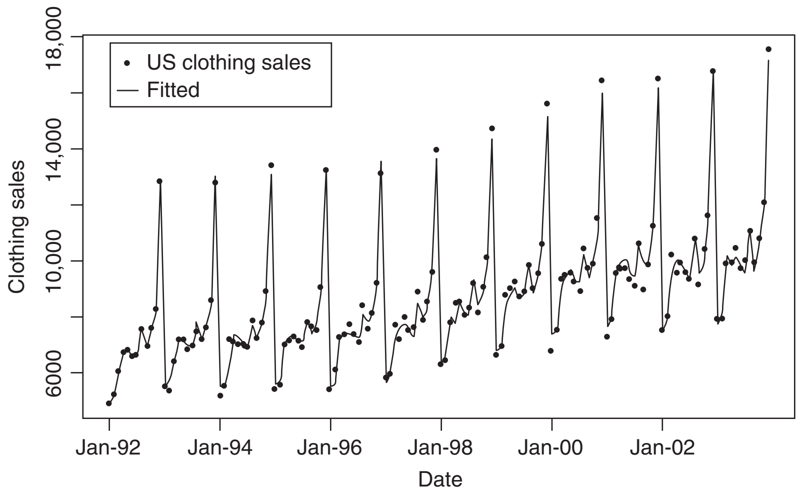
```
#ACF and PACF of the Residuals
par(mfrow=c(1,2),oma=c(0,0,0,0))
acf(res.closales.sar,lag.max=50,type="correlation",main="ACF of
the Residuals")

acf(res.closales.sar,lag.max=50,type="partial",main="PACF of the
Residuals")
```




```
#4-in-1 plot of the residuals
par(mfrow=c(2,2),oma=c(0,0,0,0))
qqnorm(res.closales.sar,datax=TRUE,pch=16,xlab='Residual',main='')
qqline(res.closales.sar,datax=TRUE)
plot(fit.closales.sar,res.closales.sar,pch=16, xlab='Fitted
Value',ylab='Residual')
abline(h=0)
hist(res.closales.sar,col="gray",xlab='Residual',main='')
plot(res.closales.sar,type="l",xlab='Observation Order',
ylab='Residual')
points(res.closales.sar,pch=16,cex=.5)
abline(h=0)
```



```
plot(closales.data[,2],type="p",pch=16,cex=.5,xlab='Date',
ylab='Clothing Sales',xaxt='n')
axis(1, seq(1,144,24), closales.data[seq(1,144,24),1])
lines(1:144, fit.closales.sar)
legend(2,17500,c("US Clothing Sales","Fitted"), pch=c(16, NA),
lwd=c(NA,.5),cex=.55,col=c("black","black"))
```



EXERCISES

5.1 Consider the time series data shown in Chapter 4, Table E4.2.

a. Fit an appropriate ARIMA model to the first 40 observations of this time series.

b. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

c. In Exercise 4.4 you used simple exponential smoothing with $\lambda = 0.2$ to smooth the first 40 time periods of this data and make forecasts of the last 10 observations. Compare the ARIMA forecasts with the exponential smoothing forecasts. How well do both of these techniques work?

5.2 Consider the time series data shown in Table E5.1.

a. Make a time series plot of the data.

b. Calculate and plot the sample autocorrelation and PACF. Is there significant autocorrelation in this time series?

TABLE E5.1 Data for Exercise 5.2

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	29	11	29	21	31	31	28	41	36
2	20	12	28	22	30	32	30	42	35
3	25	13	28	23	37	33	29	43	33
4	29	14	26	24	30	34	34	44	29
5	31	15	27	25	33	35	30	45	25
6	33	16	26	26	31	36	20	46	27
7	34	17	30	27	27	37	17	47	30
8	27	18	28	28	33	38	23	48	29
9	26	19	26	29	37	39	24	49	28
10	30	20	30	30	29	40	34	50	32

c. Identify and fit an appropriate ARIMA model to these data. Check for model adequacy.

d. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

5.3 Consider the time series data shown in Table E5.2.

a. Make a time series plot of the data.

b. Calculate and plot the sample autocorrelation and PA. Is there significant autocorrelation in this time series?

c. Identify and fit an appropriate ARIMA model to these data. Check for model adequacy.

d. Make one-step-ahead forecasts of the last 10 observations. Determine the forecast errors.

TABLE E5.2 Data for Exercise 5.3

Period	y_t	Period	y_t	Period	y_t	Period	y_t	Period	y_t
1	500	11	508	21	475	31	639	41	637
2	496	12	510	22	485	32	679	42	606
3	450	13	512	23	495	33	674	43	610
4	448	14	503	24	500	34	677	44	620
5	456	15	505	25	541	35	700	45	613
6	458	16	494	26	555	36	704	46	593
7	472	17	491	27	565	37	727	47	578
8	495	18	487	28	601	38	736	48	581
9	491	19	491	29	610	39	693	49	598
10	488	20	486	30	605	40	65	50	613

5.4 Consider the time series model

$$y_t = 200 + 0.7y_{t-1} + \varepsilon_t$$

a. Is this a stationary time series process?
b. What is the mean of the time series?
c. If the current observation is $y_{100} = 750$, would you expect the next observation to be above or below the mean?

5.5 Consider the time series model

$$y_t = 150 - 0.5y_{t-1} + \varepsilon_t$$

a. Is this a stationary time series process?
b. What is the mean of the time series?
c. If the current observation is $y_{100} = 85$, would you expect the next observation to be above or below the mean?

5.6 Consider the time series model

$$y_t = 50 + 0.8y_{t-1} - 0.15 + \varepsilon_t$$

a. Is this a stationary time series process?
b. What is the mean of the time series?
c. If the current observation is $y_{100} = 160$, would you expect the next observation to be above or below the mean?

5.7 Consider the time series model

$$y_t = 20 + \varepsilon_t + 0.2\varepsilon_{t-1}$$

a. Is this a stationary time series process?
b. Is this an invertible time series?
c. What is the mean of the time series?
d. If the current observation is $y_{100} = 23$, would you expect the next observation to be above or below the mean? Explain your answer.

5.8 Consider the time series model

$$y_t = 50 + 0.8y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t-1}$$

a. Is this a stationary time series process?

b. What is the mean of the time series?

c. If the current observation is $y_{100} = 270$, would you expect the next observation to be above or below the mean?

5.9 The data in Chapter 4, Table E4.4, exhibits a linear trend. Difference the data to remove the trend.

a. Fit an ARIMA model to the first differences.

b. Explain how this model would be used for forecasting.

5.10 Table B.1 in Appendix B contains data on the market yield on US Treasury Securities at 10-year constant maturity.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 20 observations.

c. In Exercise 4.10, you were asked to use simple exponential smoothing with $\lambda = 0.2$ to smooth the data, and to forecast the last 20 observations. Compare the ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

5.11 Table B.2 contains data on pharmaceutical product sales.

a. Fit an ARIMA model to this time series, excluding the last 10 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 10 observations.

c. In Exercise 4.12, you were asked to use simple exponential smoothing with $\lambda = 0.1$ to smooth the data, and to forecast the last 10 observations. Compare the ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

d. How would prediction intervals be obtained for the ARIMA forecasts?

5.12 Table B.3 contains data on chemical process viscosity.

a. Fit an ARIMA model to this time series, excluding the last 20 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 20 observations.

c. Show how to obtain prediction intervals for the forecasts in part b above.

5.13 Table B.4 contains data on the annual US production of blue and gorgonzola cheeses.

a. Fit an ARIMA model to this time series, excluding the last 10 observations. Investigate model adequacy. Explain how this model would be used for forecasting.

b. Forecast the last 10 observations.

c. In Exercise 4.16, you were asked to use exponential smoothing methods to smooth the data, and to forecast the last 10 observations. Compare the ARIMA and exponential smoothing forecasts. Which forecasting method do you prefer?

d. How would prediction intervals be obtained for the ARIMA forecasts?

5.14 Reconsider the blue and gorgonzola cheese data in Table B.4 and Exercise 5.13. In Exercise 4.17 you were asked to take the first difference of this data and develop a forecasting procedure based on using exponential smoothing on the first differences. Compare this procedure with the ARIMA model of Exercise 5.13.

5.15 Table B.5 shows US beverage manufacturer product shipments. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.16 Table B.6 contains data on the global mean surface air temperature anomaly. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.17 Reconsider the global mean surface air temperature anomaly data shown in Table B.6 and used in Exercise 5.16. In Exercise 4.20 you were asked to use simple exponential smoothing with the optimum value of λ to smooth the data. Compare the results with those obtained with the ARIMA model in Exercise 5.16.

5.18 Table B.7 contains daily closing stock prices for the Whole Foods Market. Develop an appropriate ARIMA model and a procedure for these data. Explain how prediction intervals would be computed.

5.19 Reconsider the Whole Foods Market data shown in Table B.7 and used in Exercise 5.18. In Exercise 4.22 you used simple exponential smoothing with the optimum value of λ to smooth the data.

Compare the results with those obtained from the ARIMA model in Exercise 5.18.

5.20 Unemployment rate data is given in Table B.8. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.21 Reconsider the unemployment rate data shown in Table B.8 and used in Exercise 5.21. In Exercise 4.24 you used simple exponential smoothing with the optimum value of λ to smooth the data. Compare the results with those obtained from the ARIMA model in Exercise 5.20.

5.22 Table B.9 contains yearly data on the international sunspot numbers. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.23 Reconsider the sunspot data shown in Table B.9 and used in Exercise 5.22.

a. In Exercise 4.26 you were asked to use simple exponential smoothing with the optimum value of λ to smooth the data, and to use an exponential smoothing procedure for trends. How do these procedures compare to the ARIMA model from Exercise 5.22? Compare the results with those obtained in Exercise 4.26.

b. Do you think that using either exponential smoothing procedure would result in better forecasts than those from the ARIMA model?

5.24 Table B.10 contains 7 years of monthly data on the number of airline miles flown in the United Kingdom. This is seasonal data.

a. Using the first 6 years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.25 Reconsider the airline mileage data in Table B.10 and used in Exercise 5.24.

a. In Exercise 4.27 you used Winters' method to develop a forecasting model using the first 6 years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from Exercise 5.24.

b. Which forecasting method would you prefer and why?

5.26 Table B.11 contains 8 years of monthly champagne sales data. This is seasonal data.

a. Using the first 7 years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.27 Reconsider the monthly champagne sales data in Table B.11 and used in Exercise 5.26.

a. In Exercise 4.29 you used Winters' method to develop a forecasting model using the first 7 years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from Exercise 5.26.

b. Which forecasting method would you prefer and why?

5.28 Montgomery et al. (1990) give 4 years of data on monthly demand for a soft drink. These data are given in Chapter 4, Table E4.5.

a. Using the first three years of data, develop an appropriate ARIMA model and a procedure for these data.

b. Explain how prediction intervals would be computed.

c. Make one-step-ahead forecasts of the last 12 months. Determine the forecast errors. How well did your procedure work in forecasting the new data?

5.29 Reconsider the soft drink demand data in Table E4.5 and used in Exercise 5.28.

a. In Exercise 4.31 you used Winters' method to develop a forecasting model using the first 7 years of data and you made forecasts for the last 12 months. Compare those forecasts with the ones you made using the ARIMA model from the previous exercise.

b. Which forecasting method would you prefer and why?

5.30 Table B.12 presents data on the hourly yield from a chemical process and the operating temperature. Consider only the yield data in this exercise. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.31 Table B.13 presents data on ice cream and frozen yogurt sales. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.32 Table B.14 presents the CO_2 readings from Mauna Loa. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.33 Table B.15 presents data on the occurrence of violent crimes. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.34 Table B.16 presents data on the US gross domestic product (GDP). Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.35 Total annual energy consumption is shown in Table B.17. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.36 Table B.18 contains data on coal production. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.37 Table B.19 contains data on the number of children 0–4 years old who drowned in Arizona. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.38 Data on tax refunds and population are shown in Table B.20. Develop an appropriate ARIMA model and a procedure for forecasting for these data. Explain how prediction intervals would be computed.

5.39 Table B.21 contains data from the US Energy Information Administration on monthly average price of electricity for the residential sector in Arizona. This data has a strong seasonal component. Use the data from 2001–2010 to develop an ARIMA model for this data. Use this model to simulate one-month-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

5.40 In Exercise 4.44 you were asked to develop a smoothing-type model for the data in Table B.21. Compare the performance of that mode with the performance of the ARIMA model from the previous exercise.

5.41 Table B.22 contains data from the Danish Energy Agency on Danish crude oil production. Develop an appropriate ARIMA model for this data. Compare this model with the smoothing models developed in Exercises 4.46 and 4.47.

5.42 Table B.23 shows Weekly data on positive laboratory test results for influenza are shown in Table B.23. Notice that these data have a number of missing values. In exercise you were asked to develop and implement a scheme to estimate the missing values. This data has a strong seasonal component. Use the data from 1997 to 2010 to develop an appropriate ARIMA model for this data. Use this model to simulate one-week-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

5.43 In Exercise 4.48 you were asked to develop a smoothing–type model for the data in Table B.23. Compare the performance of that mode with the performance of the ARIMA model from the previous exercise.

5.44 Data from the Western Regional Climate Center for the monthly mean daily solar radiation (in Langleys) at the Zion Canyon, Utah, station are shown in Table B.24. This data has a strong seasonal component. Use the data from 2003 to 2012 to develop an appropriate ARIMA model for this data. Use this model to simulate one-month-ahead forecasts for the remaining years. Calculate the forecast errors. Discuss the reasonableness of the forecasts.

5.45 In Exercise 4.50 you were asked to develop a smoothing-type model for the data in Table B.24. Compare the performance of that mode with the performance of the ARIMA model from the previous exercise.

5.46 Table B.25 contains data from the National Highway Traffic Safety Administration on motor vehicle fatalities from 1966 to 2012. This data is used by a variety of governmental and industry groups, as well as research organizations. Develop an ARIMA model for forecasting fatalities using the data from 1966 to 2006 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

5.47 Appendix Table B.26 contains data on monthly single-family residential new home sales from 1963 through 2014. Develop an ARIMA model for forecasting new home sales using the data from 1963 to 2006 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

5.48 Appendix Table B.27 contains data on the airline best on-time arrival and airport performance. The data is given by month from January 1995 through February 2013. Develop an ARIMA model for forecasting on-time arrivals using the data from 1995 to 2008 to develop the model, and then simulate one-year-ahead forecasts for the remaining years. Compute the forecasts errors. How well does this method seem to work?

5.49 Data from the US Census Bureau on monthly domestic automobile manufacturing shipments (in millions of dollars) are shown in Table B.28. Develop an ARIMA model for forecasting shipments. Note that there is some apparent seasonality in the data. Why does this seasonal behavior occur?

5.50 An ARIMA model has been fit to a time series, resulting in

$$\hat{y}_t = 25 + 0.35y_{t-1} + \varepsilon_t$$

a. Suppose that we are at time period $T = 100$ and $y_{100} = 31$. Determine forecasts for periods 101, 102, 103, … from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 33$. Revise your forecasts for periods 102, 103, … using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 2$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.51 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 25 + 0.8y_{t-1} - 0.3y_{t-2} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$ and we know that $y_{100} = 40$ and $y_{99} = 38$. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 35$. Revise your forecasts for periods 102, 103, ... using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.52 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 25 + 0.8y_{t-1} - 0.2\varepsilon_{t-1} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the forecast for period 100 was 130 and the actual observed value was $y_{100} = 140$. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observation for time period 101 turns out to be $y_{101} = 132$. Revise your forecasts for periods 102, 103, ... using period 101 as the new origin of time.

d. If your estimate $\hat{\sigma}^2 = 1.5$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.53 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 20 + \varepsilon_t + 0.45\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the observed forecast error for period 100 was 0.5 and for period 99 we know that the observed forecast error was −0.8. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

b. What is the shape of the forecast function that evolves from this model?

c. Suppose that the observations for the next four time periods turn out to be 17.5, 21.25, 18.75, and 16.75. Revise your forecasts for periods 102, 103, ... using a rolling horizon approach.

d. If your estimate $\hat{\sigma} = 0.5$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.54 The following ARIMA model has been fit to a time series:

$$\hat{y}_t = 50 + \varepsilon_t + 0.5\varepsilon_{t-1}$$

a. Suppose that we are at the end of time period $T = 100$ and we know that the observed forecast error for period 100 was 2. Determine forecasts for periods 101, 102, 103, … from this model at origin 100.

b. What is the shape of the forecast function from this model?

c. Suppose that the observations for the next four time periods turn out to be 53, 55, 46, and 50. Revise your forecasts for periods 102, 103, … using a rolling horizon approach.

d. If your estimate $\hat{\sigma} = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

5.55 For each of the ARIMA models shown below, give the forecasting equation that evolves for lead times $\tau = 1, 2, \ldots, L$. In each case, explain the shape of the resulting forecast function over the forecast lead time.

a. AR(1)

b. AR(2)

c. MA(1)

d. MA(2)

e. ARMA(1, 1)

f. IMA(1, 1)

g. ARIMA(1, 1, 0)

5.56 Use a random number generator and generate 100 observations from the AR(1) model $y_t = 25 + 0.8y_{t-1} + \varepsilon_t$. Assume that the errors are normally and independently distributed with mean zero and variance $\sigma^2 = 1$.

a. Verify that your time series is AR(1).

b. Generate 100 observations for a $N(0, 1)$ process and add these random numbers to the 100 AR(1) observations in part a to create a new time series that is the sum of AR(1) and "white noise."

c. Find the sample autocorrelation and partial autocorrelation functions for the new time series created in part b. Can you identify the new time series?

d. Does this give you any insight about how the new time series might arise in practical settings?

5.57 Assume that you have fit the following model:

$$\hat{y}_t = y_{t-1} + 0.7\varepsilon_{t-1} + \varepsilon_t$$

a. Suppose that we are at the end of time period $T = 100$. What is the equation for forecasting the time series in period 101?

b. What does the forecast equation look like for future periods 102, 103, ... ?

c. Suppose that we know that the observed value of y_{100} was 250 and forecast error in period 100 was 12. Determine forecasts for periods 101, 102, 103, ... from this model at origin 100.

d. If your estimate $\hat{\sigma} = 1$, find a 95% prediction interval on the forecast of period 101 made at the end of period 100.

e. Show the behavior of this prediction interval for future lead times beyond period 101. Are you surprised at how wide the interval is? Does this tell you something about the reliability of forecasts from this model at long lead times?

5.58 Consider the AR(1) model $y_t = 25 + 0.75y_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 1$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.59 Consider the AR(1) model $y_t = 25 + 0.75y_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 10$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Compare the results from part b with the results from part b of Exercise 5.47. How much has changing the variance of the white noise process impacted the results?

d. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

e. Compare the results from part d with the results from part c of Exercise 5.47. How much has changing the variance of the white noise process impacted the results?

5.60 Consider the AR(2) model $y_t = 25 + 0.6y_{t-1} + 0.25y_{t-2} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 1$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.61 Consider the MA(1) model $y_t = 40 + 0.4\varepsilon_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 2$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

5.62 Consider the ARMA(1, 1) model $y_t = 50 - 0.7y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t$. Assume that the variance of the white noise process is $\sigma^2 = 2$.

a. Sketch the theoretical ACF and PACF for this model.

b. Generate 50 realizations of this AR(1) process and compute the sample ACF and PACF. Compare the sample ACF and the sample

PACF to the theoretical ACF and PACF. How similar to the theoretical values are the sample values?

c. Repeat part b using 200 realizations. How has increasing the sample size impacted the agreement between the sample and theoretical ACF and PACF? Does this give you any insight about the sample sizes required for model building, or the reliability of models built to short time series?

CHAPTER 6

TRANSFER FUNCTIONS AND INTERVENTION MODELS

He uses statistics as a drunken man uses lamp posts – For support rather than illumination

Andrew Lang, Scottish poet

6.1 INTRODUCTION

The ARIMA models discussed in the previous chapter represent a general class of models that can be used very effectively in time series modeling and forecasting problems. An implicit assumption in these models is that the conditions under which the data for the time series process is collected remain the same. If, however, these conditions change over time, ARIMA models can be improved by introducing certain inputs reflecting these changes in the process conditions. This will lead to what is known as **transfer function–noise models**. These models can be seen as regression models in Chapter 3 with serially dependent response, inputs, and the error term. The identification and the estimation of these models can be challenging. Furthermore, not all standard statistical software packages possess the capability to fit such models. So far in this book, we have used

Introduction to Time Series Analysis and Forecasting, Second Edition.
Douglas C. Montgomery, Cheryl L. Jennings and Murat Kulahci.

the Minitab and JMP software packages to illustrate time series model fitting. However, Minitab (version 16) lacks the capability of fitting transfer function–noise models. Therefore for Chapters 6 and 7, we will use JMP and R instead.

6.2 TRANSFER FUNCTION MODELS

In Section 5.2, we discussed the **linear filter** and defined it as

$$\begin{aligned} y_t = L(x_t) &= \sum_{i=-\infty}^{+\infty} v_i x_{t-i} \\ &= v(B)x_t, \end{aligned} \tag{6.1}$$

where $v(B) = \sum_{i=-\infty}^{+\infty} v_i B^i$ is called the **transfer function**. Following the definition of a linear filter, Eq. (6.1) is:

1. **Time-invariant** as the coefficients $\{v_i\}$ do not depend on time.
2. **Physically realizable** if $v_i = 0$ for $i < 0$; that is, the output y_t is a linear function of the current and past values of the input:

$$\begin{aligned} y_t &= v_0 x_t + v_1 x_{t-1} + \cdots \\ &= \sum_{i=0}^{\infty} v_i x_{t-i}. \end{aligned} \tag{6.2}$$

3. **Stable** if $\sum_{i=-\infty}^{+\infty} |v_i| < \infty$.

There are two interesting special cases for the input x_t:

Impulse Response Function. If x_t is a unit impulse at time $t = 0$, that is,

$$x_t = \begin{cases} 1, & t = 0 \\ 0, & t \neq 0 \end{cases} \tag{6.3}$$

then the output y_t is

$$y_t = \sum_{i=0}^{\infty} v_i x_{t-i} = v_t \tag{6.4}$$

Therefore the coefficients v_i in Eq. (6.2) are also called the **impulse response function**.

Step Response Function. If x_t is a unit step, that is,

$$x_t = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases} \tag{6.5}$$

then the output y_t is

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} v_i x_{t-i} \\ &= \sum_{i=0}^{t} v_i, \end{aligned} \tag{6.6}$$

which is also called the **step response function**.

A generalization of the step response function is obtained when Eq. (6.5) is modified so that x_t is kept at a certain target value X after $t \geq 0$; that is,

$$x_t = \begin{cases} 0, & t < 0 \\ X, & t \geq 0. \end{cases} \tag{6.7}$$

Hence we have

$$\begin{aligned} y_t &= \sum_{i=0}^{\infty} v_i x_{t-i} \\ &= \left(\sum_{i=0}^{t} v_i \right) X \\ &= gX, \end{aligned} \tag{6.8}$$

where g is called the **steady-state gain**.

A more realistic representation of the response is obtained by adding a noise or disturbance term to Eq. (6.2) to account for unanticipated and/or ignored factors that may have an effect on the response as well. Hence the "additive" model representation of the dynamic systems is given as

$$y_t = v(B)x_t + N_t, \tag{6.9}$$

where N_t represents the unobservable noise process. In Eq. (6.9), x_t and N_t are assumed to be independent. The model representation in Eq. (6.9) is also called the **transfer function–noise model**.

Since the noise process is unobservable, the predictions of the response can be made by estimating the impulse response function $\{v_t\}$. Similar to our discussion about the estimation of the coefficients in Wold's decomposition theorem in Chapter 5, attempting to estimate the infinitely many coefficients in $\{v_t\}$ is a futile exercise. Therefore also parallel to the arguments we made in Chapter 5, we will make assumptions about these infinitely many coefficients to be able to represent them with only a handful of parameters. Following the derivations we had for the ARMA models, we will assume that the coefficients in $\{v_t\}$ have a structure and can be represented as

$$\begin{aligned} v(B) &= \sum_{i=0}^{\infty} v_i B^i = \frac{w(B)}{\delta(B)} \\ &= \frac{w_0 - w_1 B - \cdots - w_s B^s}{1 - \delta_1 B - \cdots - \delta_r B^r} \end{aligned} \tag{6.10}$$

The interpretation of Eq. (6.10) is quite similar to the one we had for ARMA models; the denominator summarizes the infinitely many coefficients with a certain structure determined by $\{\delta_i\}$ as in the AR part of the ARMA model and the numerator represents the adjustment we may like to make to the strictly structured infinitely many coefficients as in the MA part of the ARMA model.

So the transfer function–noise model in Eq. (6.9) can be rewritten as

$$y_t = \frac{w(B)}{\delta(B)} x_t + N_t,$$

where $w(B)/\delta(B) = \delta(B)^{-1} w(B) = \sum_{i=0}^{+\infty} v_i B^i$. For some processes, there may also be a **delay** before a change in the input x_t shows its effect on the response y_t. If we assume that there is b time units of delay between the response and the input, a more general representation for the transfer function–noise models can be obtained as

$$\begin{aligned} y_t &= \frac{w(B)}{\delta(B)} x_{t-b} + N_t \\ &= \frac{w(B)}{\delta(B)} B^b x_t + N_t \\ &= v(B) x_t + N_t, \end{aligned} \tag{6.11}$$