18

Factor Models and Principal Components

18.1 Dimension Reduction

High-dimensional data can be challenging to analyze. They are difficult to visualize, need extensive computer resources, and often require special statistical methodology. Fortunately, in many practical applications, high-dimensional data have most of their variation in a lower-dimensional space that can be found using *dimension reduction techniques*. There are many methods designed for dimension reduction, and in this chapter we will study two closely related techniques, *factor analysis* and *principal components analysis*, often called *PCA*.

PCA finds structure in the covariance or correlation matrix and uses this structure to locate low-dimensional subspaces containing most of the variation in the data.

Factor analysis explains returns with a smaller number of fundamental variables called *factors* or *risk factors*. Factor analysis models can be classified by the types of variables used as factors, macroeconomic or fundamental, and by the estimation technique, time series regression, cross-sectional regression, or statistical factor analysis.

18.2 Principal Components Analysis

PCA starts with a sample $\boldsymbol{Y}_i = (Y_{i,1}, \ldots, Y_{i,d})$, $i = 1, \ldots, n$, of d-dimensional random vectors with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. One goal of PCA is finding "structure" in $\boldsymbol{\Sigma}$.

We will start with a simple example that illustrates the main idea. Suppose that $\boldsymbol{Y}_i = \boldsymbol{\mu} + W_i \boldsymbol{o}$, where $W_1, \ldots, W_n$ are i.i.d. mean-zero random variables and $\boldsymbol{o}$ is some fixed vector, which can be taken to have norm 1. The $\boldsymbol{Y}_i$ lie on

D. Ruppert, D.S. Matteson, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics,
DOI 10.1007/978-1-4939-2614-5_18

the line that passes through $\boldsymbol{\mu}$ and is in the direction given by $\boldsymbol{o}$, so that all variation among the mean-centered vectors $\boldsymbol{Y}_i - \boldsymbol{\mu}$ is in the one-dimensional space spanned by $\boldsymbol{o}$. Also, the covariance matrix of $\boldsymbol{Y}_i$ is

$$\boldsymbol{\Sigma} = E\{W_i^2 \boldsymbol{o}\boldsymbol{o}^\mathsf{T}\} = \sigma_W^2 \boldsymbol{o}\boldsymbol{o}^\mathsf{T}.$$

The vector $\boldsymbol{o}$ is called the first principal axis of $\boldsymbol{\Sigma}$ and is the only eigenvector of $\boldsymbol{\Sigma}$ with a nonzero eigenvalue, so $\boldsymbol{o}$ can be estimated by an eigen-decomposition (Appendix A.20) of the estimated covariance matrix.

A slightly more realistic situation is where $\boldsymbol{Y}_i = \boldsymbol{\mu} + W_i \boldsymbol{o} + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i$ is a random vector uncorrelated with W_i and having a "small" covariance matrix. Then most of the variation among the $\boldsymbol{Y}_i - \boldsymbol{\mu}$ vectors is in the space spanned by $\boldsymbol{o}$, but there is small variation in other directions due to $\boldsymbol{\epsilon}_i$. Having looked at some simple special cases, we now turn to the general case.

PCA can be applied to either the sample covariance matrix or the correlation matrix. We will use $\boldsymbol{\Sigma}$ to represent whichever matrix is chosen. The correlation matrix is, of course, the covariance matrix of the standardized variables, so the choice between the two matrices is really a decision whether or not to standardize the variables before PCA. This issue will be addressed later. Even if the data have not been standardized, to keep notation simple, we assume that the mean $\overline{\boldsymbol{Y}}$ has been subtracted from each $\boldsymbol{Y}_i$. By (A.50),

$$\boldsymbol{\Sigma} = \boldsymbol{O}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_d)\,\boldsymbol{O}^\mathsf{T}, \tag{18.1}$$

where $\boldsymbol{O}$ is an orthogonal matrix whose columns $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_d$ are the eigenvectors of $\boldsymbol{\Sigma}$ and $\lambda_1 > \ldots > \lambda_d$ are the corresponding eigenvalues. The columns of $\boldsymbol{O}$ have been arranged so that the eigenvalues are ordered from largest to smallest. This is not essential, but it is convenient. We also assume no ties among the eigenvalues, which almost certainly will be true in actual applications.

A *normed linear combination* of $\boldsymbol{Y}_i$ (either standardized or not) is of the form $\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{Y}_i = \sum_{j=1}^p \alpha_j Y_{i,j}$, where $\|\boldsymbol{\alpha}\| = \sqrt{\sum_{j=1}^p \alpha_i^2} = 1$. The first principal component is the normed linear combination with the greatest variance. The variation in the direction $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is any fixed vector with norm 1, is

$$\mathrm{Var}(\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{Y}_i) = \boldsymbol{\alpha}^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{\alpha}. \tag{18.2}$$

The first principal component maximizes (18.2) over α. The maximizer is $\boldsymbol{\alpha} = \boldsymbol{o}_1$, the eigenvector corresponding to the largest eigenvalue, and is called the first principal axis. The projections $\boldsymbol{o}_1^\mathsf{T}\boldsymbol{Y}_i$, $i = 1, \ldots, n$, onto this vector are called the first principal component or principal component scores. Requiring that the norm of $\boldsymbol{\alpha}$ be fixed is essential, because otherwise (18.2) is unbounded and there is no maximizer.

After the first principal component has been found, one searches for the direction of maximum variation perpendicular to the first principal axis (eigenvector). This means maximizing (18.2) subject to $\|\boldsymbol{\alpha}\| = 1$ and $\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{o}_1 = 0$.

The maximizer, called the second principal axis, is $\boldsymbol{o}_2$, and the second principal component is the set of projections $\boldsymbol{o}_2^\mathsf{T}\boldsymbol{Y}_i$, $i = 1, \ldots, n$, onto this axis. The reader can probably see where we are going. The third principal component maximizes (18.2) subject to $\|\boldsymbol{\alpha}\| = 1$, $\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{o}_1 = 0$, and $\boldsymbol{\alpha}^\mathsf{T}\boldsymbol{o}_2 = 0$ and is $\boldsymbol{o}_3^\mathsf{T}\boldsymbol{Y}_i$, and so forth, so that $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_d$ are the principal axes and the set of projections $\boldsymbol{o}_j^\mathsf{T}\boldsymbol{Y}_i$, $i = 1, \ldots, n$, onto the jth eigenvector is the jth principal component. Moreover,

$$\lambda_i = \boldsymbol{o}_i^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{o}_i$$

is the variance of the ith principal component, $\lambda_i/(\lambda_1 + \cdots + \lambda_d)$ is the proportion of the variance due to this principal component, and $(\lambda_1 + \cdots + \lambda_i)/(\lambda_1 + \cdots + \lambda_d)$ is the proportion of the variance due to the first i principal components. The principal components are mutually uncorrelated since for $j \neq k$ we have

$$\mathrm{Cov}(\boldsymbol{o}_j^\mathsf{T}\boldsymbol{Y}_i, \boldsymbol{o}_k^\mathsf{T}\boldsymbol{Y}_i) = \boldsymbol{o}_j^\mathsf{T}\boldsymbol{\Sigma}\boldsymbol{o}_k = 0$$

by (A.52).

Let

$$\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1^\mathsf{T} \\ \vdots \\ \boldsymbol{Y}_n^\mathsf{T} \end{pmatrix}$$

be the original data and let

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{o}_1^\mathsf{T}\boldsymbol{Y}_1 & \cdots & \boldsymbol{o}_d^\mathsf{T}\boldsymbol{Y}_1 \\ \vdots & \ddots & \vdots \\ \boldsymbol{o}_1^\mathsf{T}\boldsymbol{Y}_n & \cdots & \boldsymbol{o}_d^\mathsf{T}\boldsymbol{Y}_n \end{pmatrix}$$

be the matrix of principal components. Then

$$\boldsymbol{S} = \boldsymbol{Y}\boldsymbol{O}.$$

Postmultiplication of $\boldsymbol{Y}$ by $\boldsymbol{O}$ to obtain $\boldsymbol{S}$ is an orthogonal rotation of the data. For this reason, the eigenvectors are sometimes called the *rotations*, e.g., in output from R's `pca()` function.

In many applications, the first few principal components, such as, the first three to five, account for almost all of the variation, and, for most purposes, one can work solely with these principal components and discard the rest. This can be a sizable reduction in dimension. See Example 18.2 for an illustration.

So far, we have left unanswered the question of how one should decide between working with the original or the standardized variables. If the components of $\boldsymbol{Y}_i$ are comparable, e.g., are all daily returns on equities or all are yields on bonds, then working with the original variables should cause no problems. However, if the variables are not comparable, e.g., one is an unemployment rate and another is the GDP in dollars, then some variables may be many orders of magnitude larger than the others. In such cases, the large

variables could completely dominate the PCA, so that the first principal component is in the direction of the variable with the largest standard deviation. To eliminate this problem, one should standardize the variables.

Example 18.1. PCA with unstandardized and standardized variables

As a simple illustration of the difference between using standardized and unstandardized variables, suppose there are two variables ($d = 2$) with a correlation of 0.9. Then the correlation matrix is

$$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$$

with normalized eigenvectors (0.71, 0.71) and (−0.71, 0.71)[1] and eigenvalues 1.9 and 0.1. Most of the variation is in the direction (1, 1), which is consistent with the high correlation between the two variables.

However, suppose that the first variable has variance 1,000,000 and the second has variance 1. The covariance matrix is

$$\begin{pmatrix} 1{,}000{,}000 & 900 \\ 900 & 1 \end{pmatrix},$$

which has eigenvectors, after rounding, equal to (1.0000, 0.0009) and (−0.0009,1) and eigenvalues 1,000,000 and 0.19. The first variable dominates the principal components analysis based on the covariance matrix. This principal components analysis does correctly show that almost all of the variation is in the first variable, but this is true only with the original units. Suppose that variable 1 had been in dollars and is now converted to millions of dollars. Then its variance is equal to 10^{-6}, so that the principal components analysis using the covariance matrix will now show most of the variation to be due to variable 2. In contrast, principal components analysis based on the correlation matrix does not change as the variables' units change. □

Example 18.2. Principal components analysis of yield curves

This example uses yields on Treasury bonds at 11 maturities, $T = 1$, 3, and 6 months and 1, 2, 3, 5, 7, 10, 20, and 30 years. Daily yields were taken from a U.S. Treasury website for the time period January 2, 1990, to October 31, 2008, A subset of these data was used in Example 15.1. The yield curves are shown in Fig. 18.1a for three different dates. Notice that the yield curves can have a variety of shapes. In this example, we will use PCA to study how the curves change from day to day.

To analyze daily changes in yields, all 11 time series were differenced. Daily yields were missing from some values of T because, for example to quote the

[1] The normalized eigenvalues are determined only up to sign so they could multiplied by −1 to become $(-0.71, -0.71)$ and $(0.71, -0.71)$.

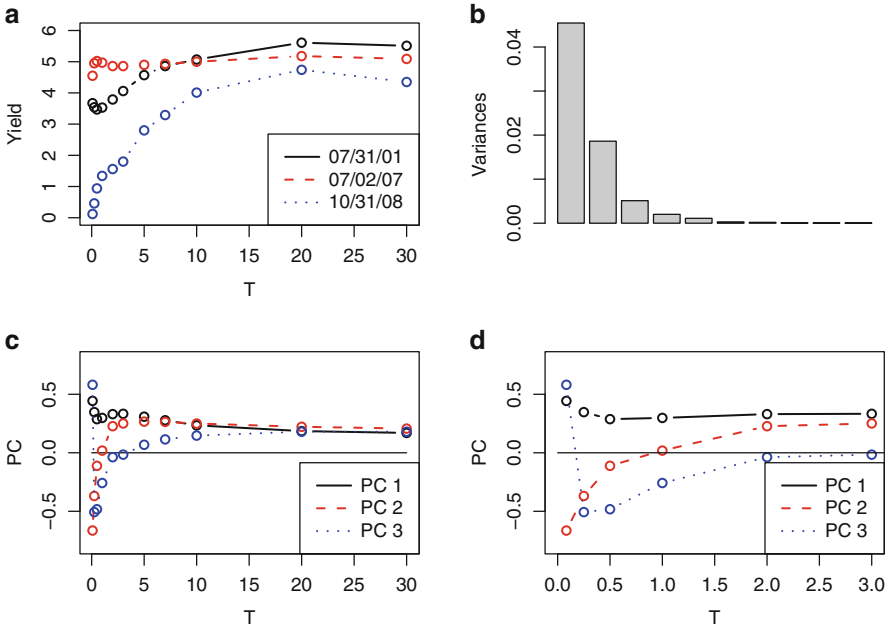

Fig. 18.1. *(a) Treasury yields on three dates. (b) Scree plot for the changes in Treasury yields. Note that the first three principal components have most of the variation, and the first five have virtually all of it. (c) The first three eigenvectors for changes in the Treasury yields. (d) The first three eigenvectors for changes in the Treasury yields in the range* $0 \leq T \leq 3$.

website, "Treasury discontinued the 20-year constant maturity series at the end of calendar year 1986 and reinstated that series on October 1, 1993." Differencing caused a few additional days to have missing values. In the analysis, all days with missing values of the differenced data were omitted. This left 819 days of data starting on July 31, 2001, when the one-month series started and ending on October 31, 2008, with the exclusion of the period February 19, 2002 to February 2, 2006 when the 30-year Treasury was discontinued. One could use much longer series by not including the one-month and 30-year series.

The covariance matrix, not the correlation matrix, was used, because in this example the variables are comparable and in the same units.

First, we will look at the 11 eigenvalues using R's function `prcomp()`. The code is:

```
datNoOmit = read.table("treasury_yields.txt", header = TRUE)
diffdatNoOmit = diff(as.matrix(datNoOmit[ , 2:12]))
dat = na.omit(datNoOmit)
diffdat = na.omit(diffdatNoOmit)
n = dim(diffdat)[1]
```

```
options(digits = 5)
pca = prcomp(diffdat)
summary(pca)
```

The results are:

```
Importance of components:
                          PC1  PC2   PC3   PC4   PC5    PC6
Standard deviation     0.21 0.14 0.071 0.045 0.033 0.0173
Proportion of Variance 0.62 0.25 0.070 0.028 0.015 0.0041
Cumulative Proportion  0.62 0.88 0.946 0.974 0.989 0.9932

PC7    PC8    PC9    PC10    PC11
0.0140 0.0108 0.0092 0.00789 0.00610
0.0027 0.0016 0.0012 0.00085 0.00051
0.9959 0.9975 0.9986 0.99949 1.00000
```

The first row gives the values of $\sqrt{\lambda_i}$, the second row the values of $\lambda_i/(\lambda_1 + \cdots + \lambda_d)$, and the third row the values of $(\lambda_1 + \cdots + \lambda_i)/(\lambda_1 + \cdots + \lambda_d)$ for $i = 1, \ldots, 11$. One can see, for example, that the standard deviation of the first principal component is 0.21 and represents 62 % of the total variance. Also, the first three principal components have 94.6 % of the variation, and this increases to 97.4 % for the first four principal components and to 98.9 % for the first five. The variances (the squares of the first row) are plotted in Fig. 18.1b. This type of plot is called a "scree plot" since it looks like scree, fallen rocks that have accumulated at the base of a mountain.

We will concentrate on the first three principal components since approximately 95 % of the variation in the changes in yields is in the space they span. The eigenvectors, labeled "PC," are plotted in Fig. 18.1c and d, the latter showing detail in the range $T \leq 3$. The eigenvectors have interesting interpretations. The first, $\boldsymbol{o}_1$, has all positive values.[2] A change in this direction either increases all yields or decreases all yields, and by roughly the same amounts. One could call such changes "parallel shifts" of the yield curve, though they are only approximately parallel. These shifts are shown in Fig. 18.2a, where the mean yield curve is shown as a solid black line, the mean plus $\boldsymbol{o}_1$ is a dashed red line, and the mean minus $\boldsymbol{o}_1$ is a dashed blue line. Only the range $T \leq 7$ is shown, since the curves change less after this point. Since the standard deviation of the first principal component is only 0.21, a ± 1 shift in a single day is huge and is used only for better graphical presentation.

[2] As mentioned previously, the eigenvectors are determined only up to a sign reversal, since multiplication by -1 would not change the spanned space or the norm. Thus, we could instead say the eigenvector has only negative values, but this would not change the interpretation.

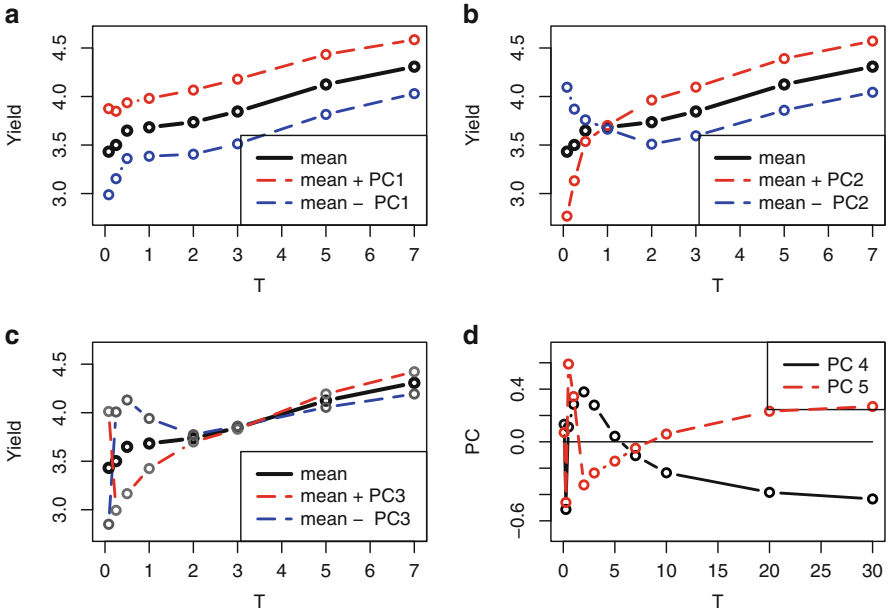

Fig. 18.2. *(a) The mean yield curve plus and minus the first eigenvector. (b) The mean yield curve plus and minus the second eigenvector. (c) The mean yield curve plus and minus the third eigenvector. (d) The fourth and fifth eigenvectors for changes in the Treasury yields.*

The graph of $\boldsymbol{o}_2$ is everywhere decreasing[3] and changes in this direction either increase or decrease the slope of the yield curve. The result is that a graph of the mean plus or minus PC2 will cross the graph of the mean curve at approximately $T = 1$, where $\boldsymbol{o}_2$ equals zero; see Fig. 18.2b.

The graph of $\boldsymbol{o}_3$ is first decreasing and then increasing, and the changes in this direction either increase or decrease the convexity of the yield curve. The result is that a graph of the mean plus or minus PC3 will cross the graph of the mean curve twice; see Fig. 18.2c. It is worth repeating a point just made in connection with PC1, since it is even more important here. The standard deviations in the directions of PC2 and PC3 are only 0.14 and 0.071, respectively, so observed changes in these directions will be much smaller than those shown in Fig. 18.2b and c. Moreover, parallel shifts will be larger than changes in slope, which will be larger than changes in convexity.

Figure 18.2d plots the fourth and fifth eigenvectors. The patterns in their graphs are complex and do not have easy interpretations. Fortunately, the variation in the space they span is too small to be of much importance.

[3] The graph would, of course, be everywhere increasing if $\boldsymbol{o}_2$ were multiplied by -1.

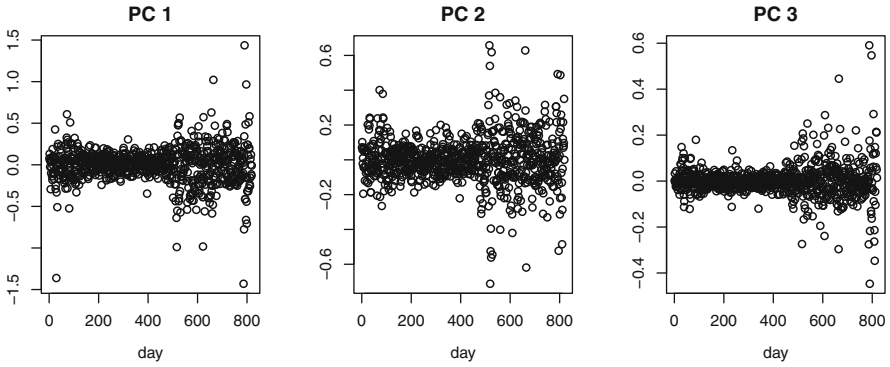

Fig. 18.3. *Time series plots of the first three principal components of the Treasury yields. There are 819 days of data, but they are not consecutive because of missing data; see text.*

A bond portfolio manager would be interested in the behavior of the yield changes over time. Time series analysis based on the changes in the 11 yields could be useful, but a better approach would be to use the first three principal components. Their time series and auto- and cross-correlation plots are shown in Figs. 18.3 and 18.4, respectively. The latter shows moderate short-term auto-correlations which could be modeled with an ARMA process, though the correlation is small enough that it might be ignored. Notice that the lag-0 cross-correlations are zero; this is not a coincidence but rather is due to the way the principal components are defined. They are defined to be uncorrelated with each other, so their lag-0 correlations are exactly zero. Cross-correlations at nonzero lags are not zero, but in this example they are small. The practical implication is that parallel shifts, changes in slopes, and changes in convexity are nearly uncorrelated and could be analyzed separately. The time series plots show substantial volatility clustering which could be modeled using the GARCH models of Chap. 14. □

Example 18.3. Principal components analysis of equity funds

This example uses the data set `equityFunds.csv`. The variables are daily returns from January 1, 2002 to May 31, 2007 on eight equity funds: EASTEU, LATAM, CHINA, INDIA, ENERGY, MINING, GOLD, and WATER. The following code was run:

```
equityFunds = read.csv("equityFunds.csv")
pcaEq = prcomp(equityFunds[ , 2:9])
summary(pcaEq)
```

The results in this example are below and are different than those for the changes in yields, because in this example the variation is less concentrated in the first few principal components. For example, the first three principal

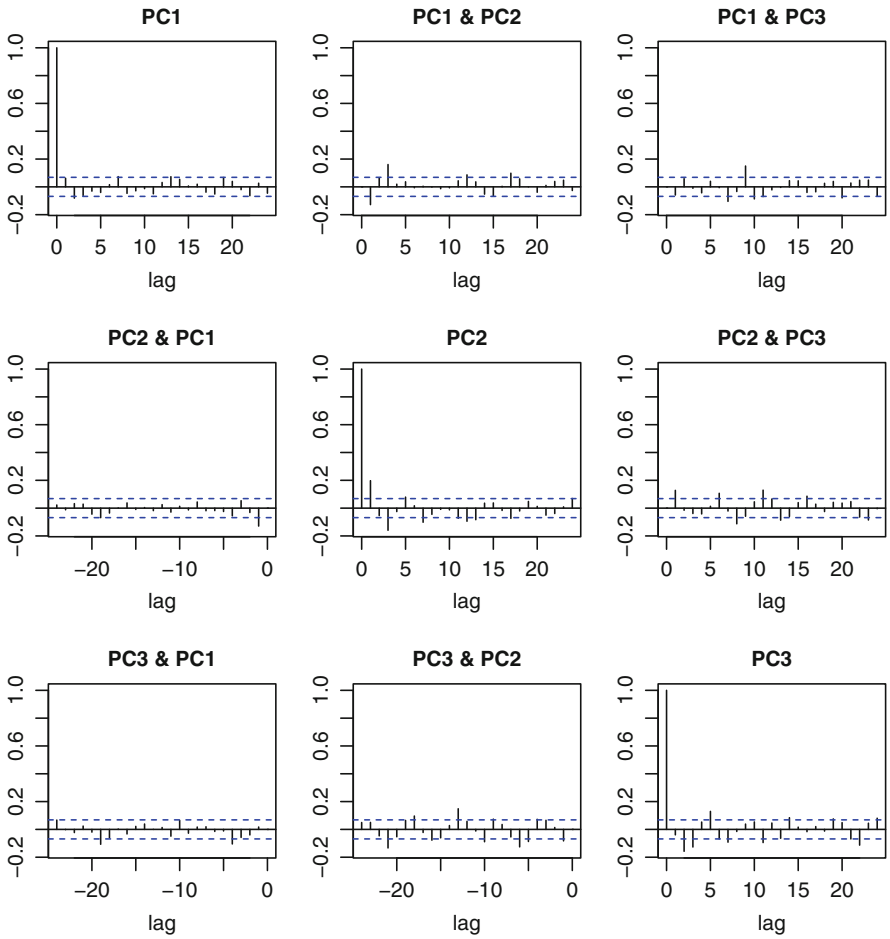

Fig. 18.4. *Sample auto- and cross-correlations of the first three principal components of the Treasury yields.*

components have only 75 % of the variance, compared to 95 % for the yield changes. For the equity funds, one needs six principal components to get 95 %. A scree plot is shown in Fig. 18.5a.

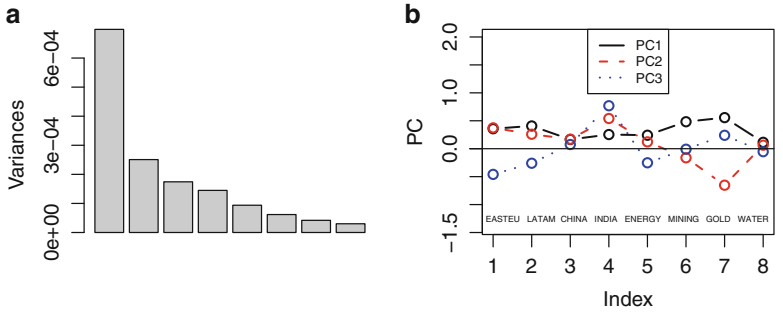

Fig. 18.5. *(a) Scree plot for the Equity Funds example. (b) The first three eigenvectors for the Equity Funds example.*

```
Importance of components:
                          PC1    PC2    PC3    PC4     PC5
Standard deviation     0.026 0.016 0.013 0.012 0.0097
Proportion of Variance 0.467 0.168 0.117 0.097 0.0627
Cumulative Proportion  0.467 0.635 0.751 0.848 0.9107

   PC6     PC7     PC8
0.0079 0.0065 0.0055
0.0413 0.0280 0.0201
0.9520 0.9799 1.0000
```

The first three eigenvectors are plotted in Fig. 18.5b. The first eigenvector has only positive values, and returns in this direction are either positive for all of the funds or negative for all of them. The second eigenvector is negative for mining and gold (funds 6 and 7) and positive for the other funds. Variation along this eigenvector has mining and gold moving in the opposite direction of the other funds. Gold and mining stock moving counter to the rest of the stock market is a common occurrence and, in fact, these types of stock often have negative betas, so it is not surprising that the second principal component has 17 % of the variation. The third principal component is less easy to interpret, but its loading on India (fund 4) is higher than on the other funds, which might indicate that there is something different about Indian equities. □

Example 18.4. Principal components analysis of the Dow Jones 30

As a further example, we will use returns on the 30 stocks on the Dow Jones average. The data are in the data set `DowJone30.csv` and cover the period from January 2, 1991 to January 2, 2002 The first five principal components have over 97 % of the variation:

```
Importance of components:
                          PC1    PC2   PC3    PC4    PC5
Standard deviation     88.53 24.967 13.44 10.602 8.2165
Proportion of Variance  0.87  0.069  0.02  0.012 0.0075
Cumulative Proportion   0.87  0.934  0.95  0.967 0.9743
```

In contrast to the analysis of the equity funds where six principal components were needed to obtain 95 % of the variance, here the first three principal components have over 95 % of the variance. Why are the Dow Jones stocks behaving differently compared to the equity funds? The Dow Jones stocks are similar to each other since they are all large companies in the United States. Thus, we can expect that their returns will be highly correlated with each other and a few principal components will explain most of the variation. □

18.3 Factor Models

A factor model for excess equity returns is

$$R_{j,t} = \beta_{0,j} + \beta_{1,j}F_{1,t} + \cdots + \beta_{p,j}F_{p,t} + \epsilon_{j,t}, \tag{18.3}$$

where $R_{j,t}$ is either the return or the excess return on the jth asset at time t, $F_{1,t}, \ldots, F_{p,t}$ are variables, called *factors* or *risk factors*, that represent the "state of the financial markets and world economy" at time t, and $\epsilon_{1,t}, \ldots, \epsilon_{n,t}$ are uncorrelated, mean-zero random variables called the *unique risks* of the individual stocks. The assumption that unique risks are uncorrelated means that all cross-correlation between the returns is due to the factors. Notice that the factors do not depend on j since they are common to all returns. The parameter $\beta_{i,j}$ is called a factor loading and specifies the sensitivity of the jth return to the ith factor. Depending on the type of factor model, either the loadings, the factors, or both the factors and the loadings are unknown and must be estimated.

The CAPM is a factor model where $p = 1$ and $F_{1,t}$ is the excess return on the market portfolio. In the CAPM, the market risk factor is the only source of risk besides the unique risk of each asset. Because the market risk factor is the only risk that any two assets share, it is the sole source of correlation between asset returns. Factor models generalize the CAPM by allowing more factors than simply the market risk and the unique risk of each asset. A *factor* can be any variable thought to affect asset returns. Examples of factors include:

1. returns on the market portfolio;
2. growth rate of the GDP;
3. interest rate on short term Treasury bills or changes in this rate;
4. inflation rate or changes in this rate;
5. interest rate spreads, for example, the difference between long-term Treasury bonds and long-term corporate bonds;

6. return on some portfolio of stocks, for example, all U.S. stocks or all stocks with a high ratio of book equity to market equity — this ratio is called BE/ME in Fama and French (1992, 1995, 1996);
7. the difference between the returns on two portfolios, for example, the difference between returns on stocks with high BE/ME values and stocks with low BE/ME values.

With enough factors, most, and perhaps all, commonalities between assets should be accounted for in the model. Then the $\epsilon_{j,t}$ should represent factors truly unique to the individual assets and therefore should be uncorrelated across j (across assets), as is being assumed.

Factor models that use macroeconomic variables such as 1–5 as factors are called *macroeconomic factor models*. *Fundamental factor models* use observable asset characteristics (fundamentals) such as 6 and 7 as factors. Both types of factor models can be fit by time series regression, the topic of the next section. Fundamental factor models can also be fit by cross-sectional regression, as explained in Sect. 18.5.

18.4 Fitting Factor Models by Time Series Regression

Equation (18.3) is a regression model. If j is fixed, then it is a univariate multiple regression model, "univariate" because there is one response (the return on the jth asset) and "multiple" since there can be several predictor variables (the factors). If we combine these models across j, then we have a multivariate regression model, that is, a regression model with more than one response. Multivariate regression is used when fitting a set of returns to factors.

As discussed in Sect. 17.6, when fitting time series regression models, one should use data at the highest sampling frequency available, which is often daily or weekly, though only monthly data were available for the next example.

Example 18.5. A macroeconomic factor model

The efficient market hypothesis implies that stock prices change because of new information. Although there is considerable debate about the extent to which markets are efficient, one still can expect that stock returns will be influenced by unpredictable changes in macroeconomic variables. Accordingly, the factors in a macroeconomic model are not the macroeconomic variables themselves, but rather the residuals when changes in the macroeconomic variables are predicted from past data by a time series model, such as, a multivariate AR model.

In this example, we look at a subset of a case study that has been presented by other authors; see the bibliographical notes in Sect. 18.7. The macroeconomic variables in this example are changes in the logs of CPI (Consumer

Price Index) and IP (Industrial Production). The changes in these series have been analyzed before in Examples 12.10, 12.11, and 13.10 and in that last example a bivariate AR model was fit. It was found that the AR(5) model minimized AIC, but the AR(1) had an AIC value nearly as small as the AR(5) model.

In this example, we will use the residuals from the AR(5) model as the factors. Monthly returns on nine stocks were taken from the `berndtInvest.csv` data set. The returns are from January 1978 to December 1987. The CPI and IP series from July 1977 to December 1987 were used, but the month of July 1977 was lost through differencing. This left enough data (the five months August 1977 to December 1977) for forecasting CPI and IP beginning January 1978 when the return series started.

R^2 and the slopes for the regressions of the stock returns on the CPI residuals and the IP residuals are plotted in Fig. 18.6 for each of the 9 stocks. Note that the R^2-values are very small, so the macroeconomic factors have little explanatory power. The problem of low explanatory power is common with macroeconomic factor models and has been noticed by other authors. For this reason, fundamental factor models are more widely used than macroeconomic models. □

18.4.1 Fama and French Three-Factor Model

Fama and French (1995) have developed a fundamental factor model with three risk factors, the first being the excess return of the market portfolio, which is the sole factor in the CAPM. The second risk factor, which is called small minus big (SMB), is the difference in returns on a portfolio of small stocks and a portfolio of large stocks. Here "small" and "big" refer to the size of the *market value*, which is the share price times the number of shares outstanding. The third factor, HML (high minus low), is the difference in returns on a portfolio of high book-to-market value (BE/ME) stocks and a portfolio of low BE/ME stocks. *Book value* is the net worth of the firm according to its accounting balance sheet. Fama and French argue that most pricing anomalies that are inconsistent with the CAPM disappear in the three-factor model. Their model of the return on the jth asset for the tth holding period is

$$R_{j,t} - \mu_{f,t} = \beta_{0,j} + \beta_{1,j}(R_{M,t} - \mu_{f,t}) + \beta_{2,j}\text{SMB}_t + \beta_{3,j}\text{HML}_t + \epsilon_{j,t},$$

where SMB_t and HML_t are the values of SMB and HML and $\mu_{f,t}$ is the risk-free rate for the tth holding period. Returns on portfolios have little autocorrelation, so the returns themselves, rather than residuals from a time series model, can be used.

Notice that this model does *not* use the size or the BE/ME ratio of the jth asset to explain returns. The coefficients $\beta_{2,j}$ and $\beta_{3,j}$ are the loading of the jth asset on SMB and HML. These loadings may, but need not, be

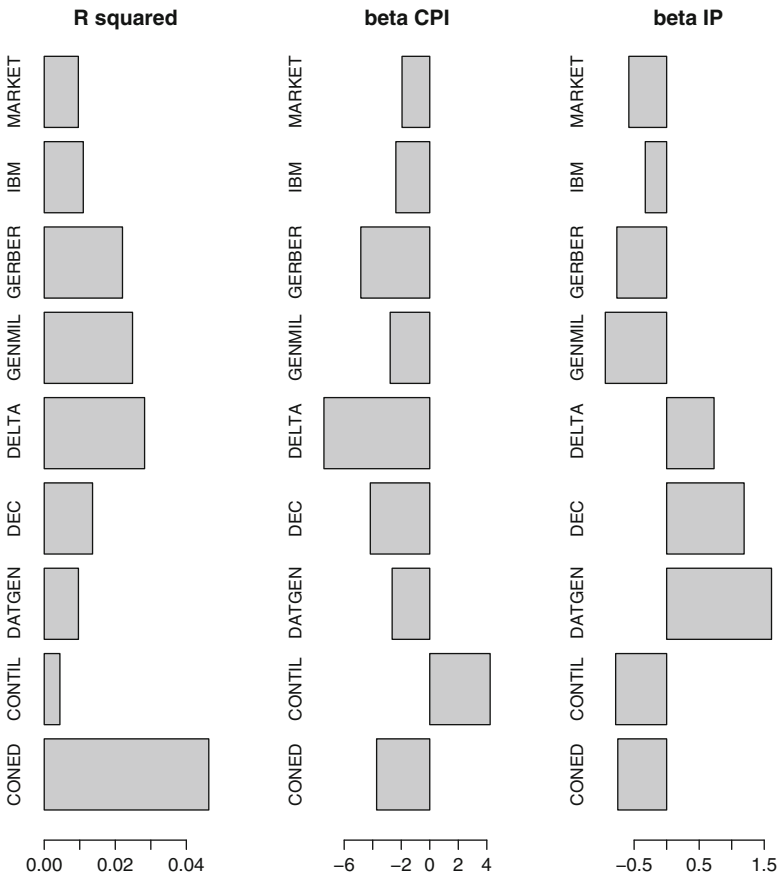

Fig. 18.6. *R^2 and slopes of regressions of stock returns on CPI residuals and IP residuals.*

related to the size and to the BE/ME ratio of the jth asset. In any event, the loadings are estimated by regression, not by measuring the size or BE/ME of the jth asset. If the loading $\beta_{2,j}$ of the jth asset on SMB is high, that might be because the jth asset is small or it might be because that asset is large but, in terms of returns, behaves similarly to small assets.

For emphasis, it is mentioned again that the factors SMB_t and HML_t do not depend on j since they are differences between returns on two fixed portfolios, not variables that are measured on the jth asset. This is true in general of the factors and loadings in model (18.3), not just the Fama–French model—only the loadings, that is, the parameters $\beta_{k,j}$, depend on the asset j. The factors are macroeconomic variables, linear combinations of returns on portfolios, or other variables that depend only on the financial markets and the economy as a whole.

There are many reasons why book and market values may differ. Book value is determined by accounting methods that do not necessarily reflect market values. Also, a stock might have a low book-to-market value because investors expect a high return on equity, which increases its market value relative to its book value. Conversely, a high book-to-market value could indicate a firm that is in trouble, which decreases its market value. A low market value relative to the book value is an indication of a stock's "cheapness," and stocks with a high market-to-book value are considered *growth stocks* for which investors are willing to pay a premium because of the promise of higher future earnings. Stocks with a low market-to-book value are called *value stocks* and investing in them is called *value investing*.

SMB and HML are the returns on portfolio that are long on one group of stocks and short on another. Such portfolios are called *hedge portfolios* since they are hedged, though perhaps not perfectly, against changes in the overall market.

Example 18.6. Fitting the Fama–French model to GE, IBM, and Mobil

This example uses two data sets. The first is `CRSPmon` in `R`'s `Ecdat` package. This is similar to the `CRSPday` data set used in previous examples except that the returns are now monthly rather than daily. There are returns on three equities, GE, IBM, and Mobil, as well as on the CRSP average, though we will not use the last one here. The returns are from January 1969 to December 1998. The second data set is the Fama–French factors and was taken from the website of Prof. Kenneth French.

Figure 18.7 is a scatterplot matrix of the GE, IBM, and Mobil excess returns and the factors. Focusing on GE, we see that, as would be expected, GE excess returns are highly correlated with the excess market returns. The GE returns are negatively related with the factor HML which would indicate that GE behaves as a growth stock, since it moves in the same direction as low BE/ME stocks and in the opposite direction of high BE/ME stocks. However, this is a false impression caused by the lack of adjustment for associations between GE excess returns and the other factors. Regression analysis will be used soon to address this problem. The two Fama–French factors are not quite hedge portfolios since SMB is positively and HML negatively related to the excess market return. However, these associations are far weaker than that between the excess returns on the stocks and the market excess returns. Moreover, SMB and HML have little association between each other, so multicollinearity is not a problem.

The three excess equity returns were regressed on the three factors using the `lm()` function in `R`. The code is:

```
FF_data = read.table("FamaFrench_mon_69_98.txt", header = TRUE)
attach(FF_data)
library("Ecdat")
```

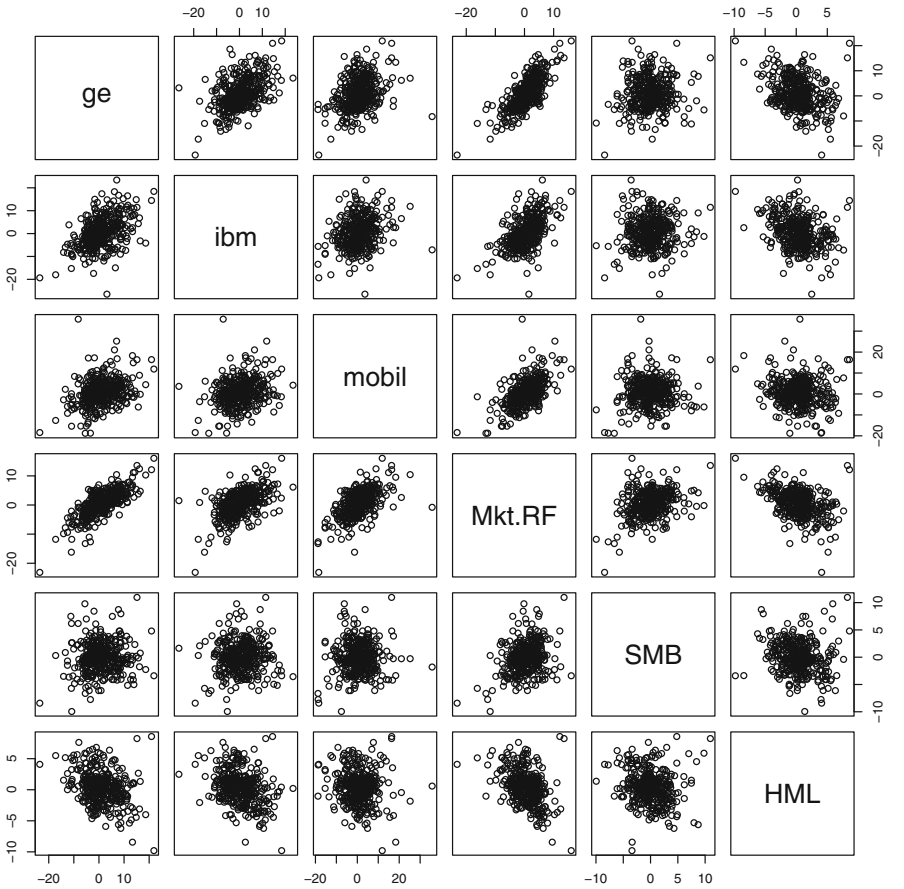

Fig. 18.7. *Scatterplot matrix of the excess returns on GE, IBM, and Mobil and the three factors in the Fama–French model.* `Mkt.RF` *is the return on the market portfolio minus the risk-free rate.*

```
library("robust")
data(CRSPmon)
ge = 100*CRSPmon[,1] - RF
ibm = 100*CRSPmon[,2] - RF
mobil = 100*CRSPmon[,3] - RF
stocks = cbind(ge, ibm, mobil)
fit = lm(cbind(ge, ibm, mobil) ~ Mkt.RF + SMB + HML)
fit
```

and the estimated coefficients are

```
Call:
lm(formula = cbind(ge, ibm, mobil) ~ Mkt.RF + SMB + HML)
```

```
Coefficients:
             ge        ibm       mobil
(Intercept)   0.3443   0.1460   0.1635
Mkt.RF        1.1407   0.8114   0.9867
SMB          -0.3719  -0.3125  -0.3753
HML           0.0095  -0.2983   0.3725
```

The coefficients of `HML` indicate that GE and Mobil are value stocks and IBM is a growth stock. Notice that GE now has a positive relationship with HML, not the negative relationship seen in Fig. 18.7, although its coefficient is close to 0. GE seems to be somewhere in between being a growth stock and a value stock.

All three equity returns have negative relationships with SMB, so, not surprisingly, they behave like large stocks.

Recall that one important assumption of the factor model is that the $\epsilon_{j,t}$ in (18.3) are uncorrelated. Violation of this assumption, that is, cross-correlations between $\epsilon_{j,t}$ and $\epsilon_{j',t}$, $j \neq j'$, will create biases when the factor model is used to estimate correlations between the equity returns, a topic explained in the next section. Lack of cross-correlation is not an assumption of the multivariate regression model and does not cause bias in the estimation of the regression coefficients or the variances of the $\epsilon_{j,t}$. The biases arise only when estimating covariances between the equity returns.

To check for cross-correlations, we will use the residuals from the multivariate regression. Their sample correlation matrix is

```
> cor(fit$residuals)
           ge    ibm mobil
ge      1.000  0.071 -0.25
ibm     0.071  1.000 -0.10
mobil -0.254 -0.102  1.00
```

The correlation between GE and Mobil is rather far from zero and is worth checking. A 95 % confidence interval for the residual correlations between GE excess returns and Mobil excess returns does not include 0, so a test would reject the null hypotheses that the true correlation is 0. The other correlations are not significantly different from 0. Because of the large negative GE–Mobil correlation, we should be careful about using the Fama–French model for estimation of the covariance matrix of the equity returns. As always, it is good practice to look at scatterplot matrices as well as correlations, since scatterplots may be outliers or nonlinear relationships affecting the correlations. Figure 18.8 contains a scatterplot matrix of the residuals. One sees that there are few outliers. although none of the outliers is really extreme, it seems worthwhile to compute robust correlations estimates and to compare them with the ordinary sample correlation matrix. Robust estimates were found using the function `covRob()` in `R`'s `robust` package. What was found is that the robust estimates are all closer to zero than the nonrobust estimates, but the robust correlation estimate for GE and Mobil is still a large negative value.

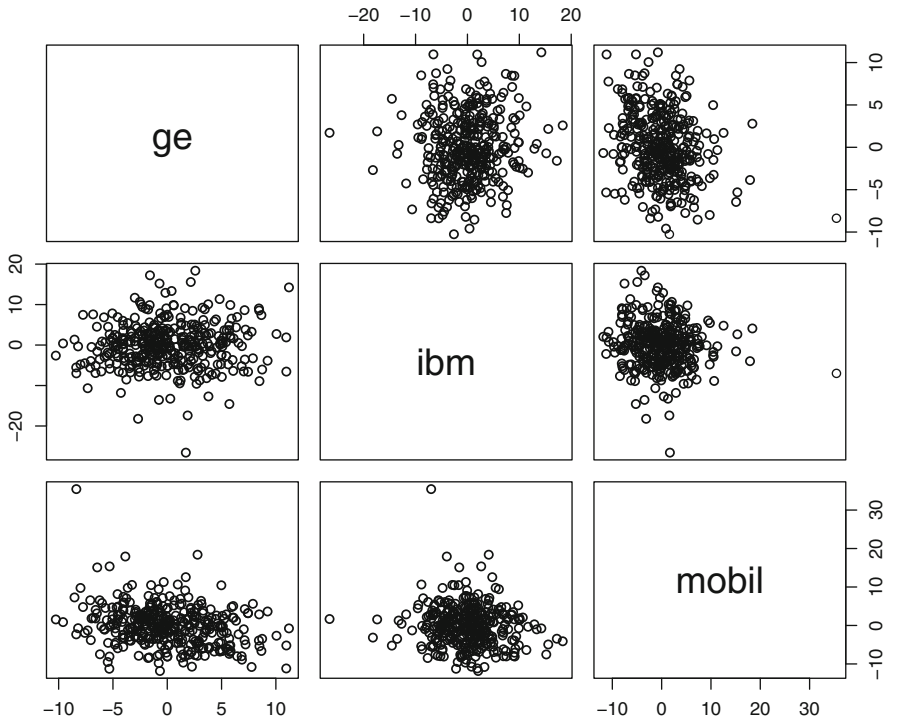

Fig. 18.8. *Scatterplot matrix of the residuals for GE, IBM, and Mobil from the Fama–French model.*

```
Call:
covRob(data = fit$residuals, corr = T)

Robust Estimate of Correlation:
           ge     ibm   mobil
ge      1.000  0.0360 -0.2479
ibm     0.036  1.0000 -0.0687
mobil -0.248 -0.0687  1.0000
```

This example is atypical of real applications because, for illustration purposes, the number of returns has been kept low, only three, whereas in portfolio management the number of returns will be larger and might be in the hundreds. □

18.4.2 Estimating Expectations and Covariances of Asset Returns

Section 17.7 discussed how the CAPM can simplify the estimation of expectations and covariances of asset returns. However, using the CAPM for this

purpose can be dangerous since the estimates depend on the validity of the CAPM. Fortunately, it is also possible to estimate return expectations and covariances using a more realistic factor model instead of the CAPM.

We start with two factors for simplicity. From (18.3), now with $p = 2$, we have

$$R_{j,t} = \beta_{0,j} + \beta_{1,j} F_{1,t} + \beta_{2,j} F_{2,t} + \epsilon_{j,t}. \tag{18.4}$$

It follows from (18.4) that

$$E(R_{j,t}) = \beta_{0,j} + \beta_{1,j} E(F_{1,t}) + \beta_{2,j} E(F_{2,t}) \tag{18.5}$$

and

$$\mathrm{Var}(R_{j,t}) = \beta_{1,j}^2 \mathrm{Var}(F_1) + \beta_{2,j}^2 \mathrm{Var}(F_2) + 2\beta_{1,j}\beta_{2,j}\mathrm{Cov}(F_1, F_2) + \sigma_{\epsilon,j}^2.$$

Also, because $R_{j,t}$ and $R_{j',t}$ are two linear combinations of the risk factors, it follows from (7.8) that for any $j \neq j'$,

$$\begin{aligned}\mathrm{Cov}(R_{j,t}, R_{j',t}) = {} & \beta_{1,j}\beta_{1,j'}\mathrm{Var}(F_1) + \beta_{2,j}\beta_{2,j'}\mathrm{Var}(F_2) \\ & + (\beta_{1,j}\beta_{2,j'} + \beta_{1,j'}\beta_{2,j})\mathrm{Cov}(F_1, F_2).\end{aligned} \tag{18.6}$$

More generally, let

$$\boldsymbol{F}_t^\mathsf{T} = (F_{1,t}, \ldots, F_{p,t}) \tag{18.7}$$

be the vector of p factors at time t and suppose that $\boldsymbol{\Sigma}_F$ is the $p \times p$ covariance matrix of $\boldsymbol{F}_t$. Define the vector of intercepts

$$\boldsymbol{\beta}_0^\mathsf{T} = (\beta_{0,1}, \ldots, \beta_{0,n})$$

and the matrix of loadings

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{1,j} & \cdots & \beta_{1,n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \beta_{p,1} & \cdots & \beta_{p,j} & \cdots & \beta_{p,n} \end{pmatrix}.$$

Also, define

$$\boldsymbol{\epsilon}^\mathsf{T} = (\epsilon_{1,t}, \ldots, \epsilon_{n,t}) \tag{18.8}$$

and let $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ be the $n \times n$ diagonal covariance matrix of $\boldsymbol{\epsilon}$:

$$\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \begin{pmatrix} \sigma_{\epsilon,1}^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\epsilon,j}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & \sigma_{\epsilon,n}^2 \end{pmatrix}.$$

Finally, let

$$\boldsymbol{R}_t^{\mathsf{T}} = (R_{1,t} \ldots, R_{n,t}) \tag{18.9}$$

be the vector of all returns at time t. Model (18.3) then can be reexpressed in matrix notation as

$$\boldsymbol{R}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{F}_t + \boldsymbol{\epsilon}_t. \tag{18.10}$$

Therefore, the $n \times n$ covariance matrix of $\boldsymbol{R}_t$ is

$$\boldsymbol{\Sigma}_R = \boldsymbol{\beta}^{\mathsf{T}} \boldsymbol{\Sigma}_F \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}. \tag{18.11}$$

In particular, if $\boldsymbol{\beta}_j = (\beta_{1,j} \quad \cdots \quad \beta_{p,j})^{\mathsf{T}}$ is the jth column of $\boldsymbol{\beta}$, then the variance of the jth return is

$$\mathrm{Var}(R_j) = \boldsymbol{\beta}_j^{\mathsf{T}} \boldsymbol{\Sigma}_F \boldsymbol{\beta}_j + \sigma_{\epsilon_j}^2, \tag{18.12}$$

and the covariance between the jth and j'th returns is

$$\mathrm{Cov}(R_j, R_j') = \boldsymbol{\beta}_j^{\mathsf{T}} \boldsymbol{\Sigma}_F \boldsymbol{\beta}_{j'}. \tag{18.13}$$

To use (18.11), (18.12) or (18.13), one needs estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_F$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. The regression coefficients are used to estimate $\boldsymbol{\beta}$, the sample covariance of the factors can be used to estimate $\boldsymbol{\Sigma}_F$, and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}$ can be the diagonal matrix of the mean residual sum of squared errors from the regressions; see equation (9.13).

Why estimate $\boldsymbol{\Sigma}_R$ via a factor model instead of simply using the sample covariance matrix? One reason is estimation accuracy. This is another example of bias–variance tradeoff. The sample covariance matrix is unbiased, but it contains $n(n+1)/2$ estimates, one for each covariance and each variance. Each of these parameters is estimated with error and when this many errors accumulate, the result can be a sizable loss of precision. In contrast, the factor model requires estimates of $n \times p$ parameters in $\boldsymbol{\beta}$, $p(p+1)/2$ parameters in $\boldsymbol{\Sigma}_F$, and n parameters in the diagonal matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, for a total of $np+n+p(p+1)/2$ parameters. Typically, n, the number of returns, is large but p, the number of factors, is much smaller, so $np+n+p(p+1)/2$ is much smaller than $n(n+1)/2$. For example, suppose there are 200 returns and 5 factors. Then $n(n+1)/2 =$ 20,100 but $np + n + p(p+1)/2$ is only 1,215. The downside of the factor model is that there will be bias in the estimate of $\boldsymbol{\Sigma}_R$ if the factor model is misspecified, especially if $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is not diagonal as the factor model assumes.

Another advantage of the factor model is expediency. Having fewer parameters to estimate is one convenience and another is ease of updating. Suppose a portfolio manager has implemented a factor model for n equities and now needs to add another equity. If the manager uses the sample covariance matrix, then the n sample covariances between the new return time series and the old ones must be computed. This requires that all n of the old time series be available. In comparison, with a factor model, the portfolio manager needs only to regress the new return time series on the factors. Only the p factor time series need to be available.

Example 18.7. Estimating the covariance matrix of GE, IBM, and Mobil excess returns

This example continues Example 18.6. Recall that the number of returns has been kept artificially low, since with more returns it would not have been possible to display the results. Therefore, this example merely illustrates the calculations and is not a typical application of factor modeling.

The estimate of $\boldsymbol{\Sigma}_F$ is the sample covariance matrix of the factors:

```
         Mkt.RF     SMB     HML
Mkt.RF 21.1507  4.2326 -5.1045
SMB     4.2326  8.1811 -1.0760
HML    -5.1045 -1.0760  7.1797
```

The estimate of $\boldsymbol{\beta}$ is the matrix of regression coefficients (without the intercepts):

```
        Mkt.RF      SMB       HML
ge     1.14071 -0.37193  0.009503
ibm    0.81145 -0.31250 -0.298302
mobil  0.98672 -0.37530  0.372520
```

The estimate of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is the diagonal matrix of residual error MS values:

```
       [,1]   [,2]   [,3]
[1,] 16.077  0.000  0.000
[2,]  0.000 31.263  0.000
[3,]  0.000  0.000 27.432
```

Therefore, the estimate of $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}_F\boldsymbol{\beta}$ is

```
           ge    ibm  mobil
ge     24.960 19.303 19.544
ibm    19.303 15.488 14.467
mobil  19.544 14.467 16.155
```

and the estimate of $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}_F\boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ is

```
           ge    ibm  mobil
ge     41.036 19.303 19.544
ibm    19.303 46.752 14.467
mobil  19.544 14.467 43.587
```

For comparison, the sample covariance matrix of the equity returns is

```
           ge    ibm  mobil
ge     40.902 20.878 14.255
ibm    20.878 46.491 11.518
mobil  14.255 11.518 43.357
```

The largest difference between the estimate of $\boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}_F\boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ and the sample covariance matrix is in the covariance between the excess returns on GE and Mobil. The reason for this large discrepancy is that the factor model assumes a zero residual correlation between these two variables, but, as we learned earlier, the data show a negative correlation of -0.25.

The code for the calculations in this example continues the code in Example 18.6. The addition code is:

```
sigF = as.matrix(var(cbind(Mkt.RF, SMB, HML)))
bbeta = as.matrix(fit$coef)
bbeta = t( bbeta[-1, ])
n = dim(CRSPmon)[1]
sigeps = (n - 1) / (n - 4) * as.matrix((var(as.matrix(fit$resid))))
sigeps = diag(as.matrix(sigeps))
sigeps = diag(sigeps, nrow = 3)
cov_equities = bbeta %*% sigF %*% t(bbeta) + sigeps
options(digits = 5)
sigF
bbeta
sigeps
bbeta %*% sigF %*% t(bbeta)
cov_equities
cov(stocks)
```

□

18.5 Cross-Sectional Factor Models

Models of the form (18.3) are *time series factor models*. They use time series data, one single asset at a time, to estimate the loadings.

As just discussed, time series factor models do not make use of variables such as dividend yields, book-to-market value, or other variables specific to the jth firm. An alternative is a *cross-sectional factor model*, which is a regression model using data from many assets but from only a single holding period. For example, suppose that R_j, $(B/M)_j$, and D_j are the return, book-to-market value, and dividend yield for the jth asset for some fixed time t. Since t is fixed, it will not be made explicit in the notation. Then a possible cross-sectional factor model is

$$R_j = \beta_0 + \beta_1(B/M)_j + \beta_2 D_j + \epsilon_j.$$

The parameters β_1 and β_2 are unknown values at time t of a book-to-market value risk factor and a dividend yield risk factor. These values are estimated by regression.

There are two fundamental differences between time series factor models and cross-sectional factor models. The first is that with a time series factor model one estimates parameters, one asset at a time, using multiple holding

periods, while in a cross-sectional model one estimates parameters, one single holding period at a time, using multiple assets. The other major difference is that in a time series factor model, the factors are directly measured and the loadings are the unknown parameters to be estimated by regression. In a cross-sectional factor model the opposite is true; the loadings are directly measured and the factor values are estimated by regression.

Example 18.8. An industry cross-sectional factor model

This example uses the `berndtInvest.csv` used in Example 18.5. This data set has monthly returns on 15 stocks over 10 years, 1978 to 1987. The 15 stocks were classified into three industries,"Tech," "Oil," and "Other," as follows:

	tech	oil	other
CITCRP	0	0	1
CONED	0	0	1
CONTIL	0	1	0
DATGEN	1	0	0
DEC	1	0	0
DELTA	0	1	0
GENMIL	0	0	1
GERBER	0	0	1
IBM	1	0	0
MOBIL	0	1	0
PANAM	0	1	0
PSNH	0	0	1
TANDY	1	0	0
TEXACO	0	1	0
WEYER	0	0	1

We used the indicator variables of "tech" and "oil" as loadings and fit the model

$$R_j = \beta_0 + \beta_1 \texttt{tech}_j + \beta_2 \texttt{oil}_j + \epsilon_j, \tag{18.14}$$

where R_j is the return on the jth stock, $\texttt{tech}_j$ equals 1 if the jth stock is a technology stock and equals 0 otherwise, and $\texttt{oil}_j$ is defined similarly. Model (18.14) was fit separately for each of the 120 months. The estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\beta}_3$ for a month were the values of the three factors for that month. The loadings were the known values of $\texttt{tech}_j$ and $\texttt{oil}_j$.

Factor 1, the values of $\widehat{\beta}_0$, can be viewed as an overall market factor, since it affects all 15 returns. Factors 2 and 3 are the technology and oil factors. For example, if the value of factor 2 is positive in any given month, then Tech stocks have better-than-market returns that month. Figure 18.9 contains time series plots of the three factor series, and Fig. 18.10 shows their auto- and

cross-correlation functions. The largest cross-correlation is between the tech and oil factors at lag 0, which indicates that above- (below-) market returns for technology stocks are associated with above (below) market returns for oil stocks.

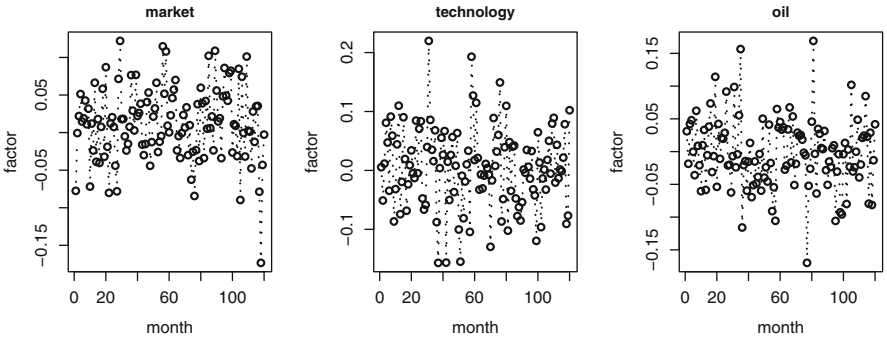

Fig. 18.9. *Time series plots of the estimated values of the three factors in the cross-sectional factor model.*

The standard deviations of the three factors are

```
market   tech    oil
 0.049  0.069  0.053
```

There are other ways of defining the factors. For example, Zivot and Wang (2006) use the model

$$R_j = \beta_1 \texttt{tech}_j + \beta_2 \texttt{oil}_j + \beta_3 \texttt{other}_j + \epsilon_j, \tag{18.15}$$

with no intercept but with $\texttt{other}_J$ as a third variable. With this model, there is no market factor but instead factors for all three industries. The model with an intercept but without `other` is equivalent to the model with `other` in place of the intercept, in the sense that the two models produce the same fitted values. □

Cross-sectional factor models are sometimes called BARRA models after BARRA, Inc., a company that has been developing cross-sectional factor models and marketing the output of their models to financial managers.

18.6 Statistical Factor Models

In a statistical factor model, neither the factor values nor the loadings are directly observable. All that is available is the sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ or, perhaps, only the sample covariance matrix. This is the same type of data available

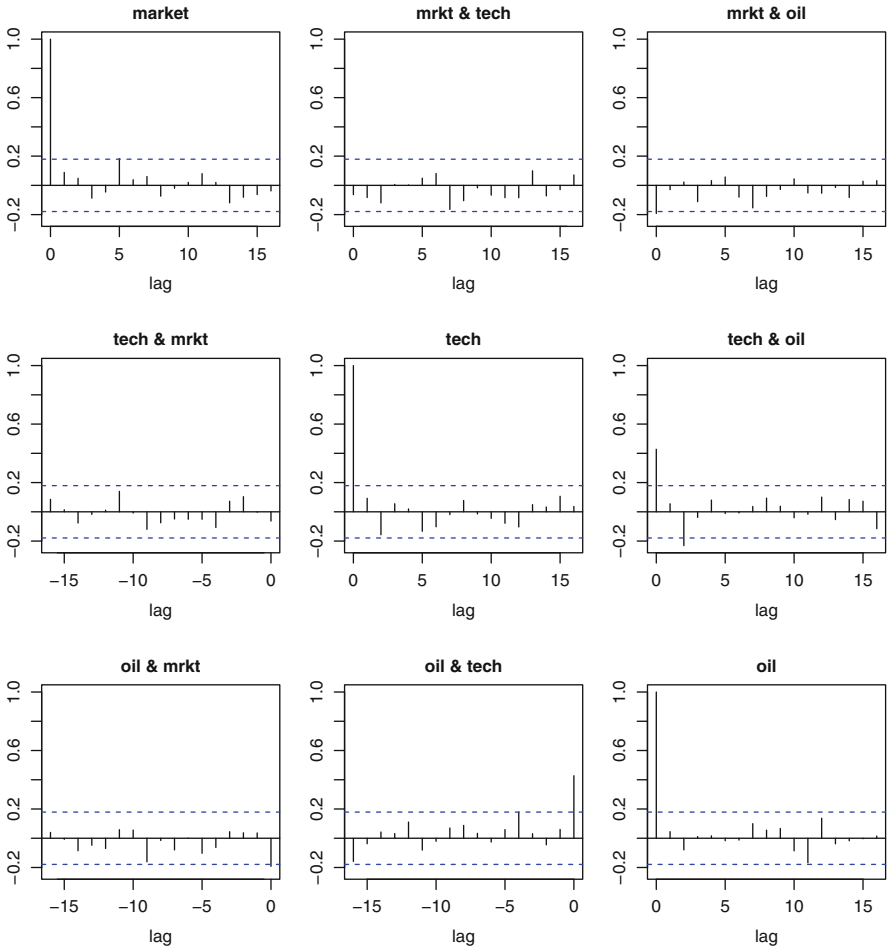

Fig. 18.10. *Auto- and cross-correlation plots of the estimated three factors in the cross-sectional factor model. Series 1–3 are the market, tech, and oil factors, respectively.*

for PCA and we will see that statistical factor analysis and PCA have some common characteristics. As with PCA, one can work with either the standardized or unstandardized variables. R's `factanal()` function automatically standardizes the variables.

We start with the multifactor model in matrix notation (18.10) and the return covariance matrix (18.11) which for convenience will be repeated as

$$\boldsymbol{R}_t = \boldsymbol{\beta}_0 + \boldsymbol{\beta}^\top \boldsymbol{F}_t + \boldsymbol{\epsilon}_t. \tag{18.16}$$

and

$$\boldsymbol{\Sigma}_R = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{\Sigma}_F \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}. \tag{18.17}$$

Here $\boldsymbol{\beta}^\mathsf{T}$ is $d \times p$ where d is the dimension of R_t and p is the number of factors.

The only component of (18.17) that can be estimated directly from the data is $\boldsymbol{\Sigma}_R$. One can use this estimate to find estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_F$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. However, it is too much to ask that all three of these matrices be identified from $\boldsymbol{\Sigma}_R$ alone. Here is the problem: Let $\boldsymbol{A}$ be any $p \times p$ invertible matrix. Then the returns vector $\boldsymbol{R}_t$ in (18.16) is unchanged if $\boldsymbol{\beta}^\mathsf{T}$ is replaced by $\boldsymbol{\beta}^\mathsf{T}\boldsymbol{A}^{-1}$ and $\boldsymbol{F}_t$ is replaced by $\boldsymbol{A}\boldsymbol{F}_t$. Therefore, the returns only determine $\boldsymbol{\beta}$ and $\boldsymbol{F}_t$ up to a nonsingular linear transformation, and consequently a set of constraints is needed to identify the parameters. The usual constraints are the factors are uncorrelated and standardized, so that

$$\boldsymbol{\Sigma}_F = \boldsymbol{I}, \tag{18.18}$$

where $\boldsymbol{I}$ is the $p \times p$ identity matrix. With these constraints, (18.17) simplifies to the statistical factor model

$$\boldsymbol{\Sigma}_R = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}. \tag{18.19}$$

However, even with this simplification, $\boldsymbol{\beta}$ is only determined up to a rotation, that is, by multiplication by an orthogonal matrix. To appreciate why this is so, let $\boldsymbol{P}$ be any orthogonal matrix, so that $\boldsymbol{P}^\mathsf{T} = \boldsymbol{P}^{-1}$. Then (18.19) is unchanged if $\boldsymbol{\beta}$ is replaced by $\boldsymbol{P}\boldsymbol{\beta}$ since

$$(\boldsymbol{P}\boldsymbol{\beta})^\mathsf{T}(\boldsymbol{P}\boldsymbol{\beta}) = \boldsymbol{\beta}^\mathsf{T}\boldsymbol{P}^\mathsf{T}\boldsymbol{P}\boldsymbol{\beta} = \boldsymbol{\beta}^\mathsf{T}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{\beta} = \boldsymbol{\beta}^\mathsf{T}\boldsymbol{\beta}.$$

Therefore, to determine $\boldsymbol{\beta}$ a further set of constraints is needed. One possible set of constraints is that $\boldsymbol{\beta}\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}^{-1}\boldsymbol{\beta}^\mathsf{T}$ is diagonal (Mardia et al., 1979, p. 258). Output from R's function `factanal()` satisfies this constraint when the argument `rotation` is set to `"none"`. If $\boldsymbol{\beta}$ is rotated as discussed in Sect. 18.6.1, then this constraint no longer holds.

If the main purpose of the statistical factor model is to estimate $\boldsymbol{\Sigma}_R$ by (18.19), then the choice of constraint is irrelevant since all constraints lead to the same product $\boldsymbol{\beta}^\mathsf{T}\boldsymbol{\beta}$. In particular, rotation of $\boldsymbol{\beta}$ does not change the estimate of $\boldsymbol{\Sigma}_R$.

It is helpful to compare three estimates of $\boldsymbol{\Sigma}_R$. The sample covariance matrix has full rank (rank $= d$) provided that $n > d$ as will be assumed here. Instead of the sample covariance matrix, one can perform PCA and estimate $\boldsymbol{\Sigma}_R$ by the sample covariance matrix of the first $p < d$ principal components. Then

$$\widehat{\boldsymbol{\Sigma}}_R = \boldsymbol{O}^\mathsf{T}\boldsymbol{O}.$$

where $\boldsymbol{O}^\mathsf{T}$ is the $d \times p$ matrix whose columns are the first d principal axes (eigenvectors) and the rank of $\widehat{\boldsymbol{\Sigma}}_R$ is only p so less than full rank. In contrast, (18.19) provides a full-rank estimate of $\boldsymbol{\Sigma}_R$ but with a simple structure, the sum of a rank p matrix and a diagonal matrix.

Example 18.9. Factor analysis of equity funds

This example continues the analysis of the equity funds data set that was used in Example 18.3 to illustrate PCA. The code for fitting a 4-factor model ($p = 4$) using `factanal()` is:

```
equityFunds = read.csv("equityFunds.csv")
fa_none = factanal(equityFunds[ , 2:9], 4, rotation = "none")
print(fa_none,cutoff = 0.1)
```

Here we specify no rotations. The output is:

```
> factanal(equityFunds[,2:9],4,rotation="none")

Call:
factanal(x = equityFunds[, 2:9], factors = 4,
      rotation = "none")

Uniquenesses:
EASTEU  LATAM  CHINA  INDIA ENERGY MINING   GOLD  WATER
 0.735  0.368  0.683  0.015  0.005  0.129  0.005  0.778

Loadings:
       Factor1 Factor2 Factor3 Factor4
EASTEU  0.387   0.169   0.293
LATAM   0.511   0.167   0.579
CHINA   0.310   0.298   0.362
INDIA   0.281   0.951
ENERGY  0.784                   0.614
MINING  0.786           0.425  -0.258
GOLD    0.798                  -0.596
WATER   0.340           0.298   0.109

               Factor1 Factor2 Factor3 Factor4
SS loadings       2.57    1.07    0.82    0.82
Proportion Var    0.32    0.13    0.10    0.10
Cumulative Var    0.32    0.46    0.56    0.66

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 17 on 2 degrees of freedom.
The p-value is 2e-04
```

The "loadings" are the estimates $\widehat{\boldsymbol{\beta}}^{\mathsf{T}}$. Since there are eight funds and four factors, the loadings are in an 8×4 matrix `fa_none$loadings`. The output above gives the sums of squares of the eight loadings for each factor. The `Proportion Var` row contains the `SS loadings` divided by 8, where 8 is the sum of the variances of the eight variables, since each variable has been standardized to have variance equal to 1.

By convention, any loading with an absolute value less than the parameter `cutoff` is not printed, and the default value of `cutoff` is 0.1.

Because all its loadings have the same sign, the first factor is an overall index of the eight funds. The second factor has large loadings on the four regional funds (EASTEU, LATAM, CHINA, INDIA) and small loadings on the four industry section funds (ENERGY, MINING, GOLD, WATER). The four regions are all emerging markets, so the second factor might be interpreted as an emerging markets factor. The fourth factor is a contrast of MINING and GOLD with ENERGY and WATER, and mimics a hedge portfolio that is long on ENERGY and WATER and short on GOLD and MINING. The third factor is less interpretable. The uniquenesses are the diagonal elements of the estimate $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}$.

The output gives a p-value for testing the null hypothesis that there are at most four factors. The p-value is small, indicating that the null hypothesis should be rejected. However, four is that maximum number of factors that can be used by `factanal()` when there are only eight returns. Should we be concerned that we are not using enough factors? Recall the important distinction between statistical and practical significance that has been emphasized elsewhere in this book. One way to assess practical significance is to see how well the factor model can reproduce the sample correlation matrix. Since `factanal()` standardizes the variables, the factor model estimate of the correlation matrix is the estimate of the covariance matrix, that, using (18.19), is

$$\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}. \tag{18.20}$$

The code to calculate this estimate is

```
B_none = fa_none$loadings[ , ]
BB_none = B_none %*% t(B_none)
D_none = diag(fa_none$unique)
Sigma_R_hat = BB_none + D_none
```

Here `B_none` is $\widehat{\boldsymbol{\beta}}^{\mathsf{T}}$ with no rotation, `BB_none` equals $\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\boldsymbol{\beta}}$ and `D_none` equals $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}$.

The difference between this estimate and the sample correlation matrix is a 8×8 matrix. We would like all of its entries to be close to 0. Unfortunately, they are not as small as we would like. There are various ways to check if a matrix this size is "small." The smallest entry is -0.063 and the largest is 0.03. These are reasonably large discrepancies between correlation matrices. Also, the eigenvalues of the difference are

```
-7.5e-02 -6.0e-03 -3.4e-15 -2.0e-15
-1.3e-15  3.0e-15  7.7e-03  7.3e-02
```

Another way to check for smallness of the difference between the two estimates is to look at the estimates of the variance of an equally weighted portfolio (of the standardized returns), which is

$$\boldsymbol{w}^{\mathsf{T}} \boldsymbol{\Sigma}_R \boldsymbol{w},$$

where $\boldsymbol{w}^{\mathsf{T}} = (1/8, \ldots, 1/8)$. These estimates are 0.37 and 0.42 using the factor model and the sample correlation matrix, respectively. The absolute difference, 0.06, is relatively large compared to either of the estimates. It is unclear whether this difference is due to a more parsimonious and accurate fit by the factor model (good) or due to bias from a lack of fit by the factor model (not good). □

18.6.1 Varimax Rotation of the Factors

As discussed earlier, the estimate of the covariance matrix is unchanged if the loadings $\boldsymbol{\beta}$ are rotated by multiplication by an orthogonal matrix. Rotation might increase the interpretability of the loadings. In some applications, it is desirable for each loading to be either close to 0 or large, so that a variable will load only on a few factors, or even on only one factor. *Varimax* rotation attempts to make each loading either small or large by maximizing the sum of the variances of the squared loadings. Varimax rotation is the default with R's `factanal()` function, but this can be changed as in Example 18.9 where no rotation was used. In finance, having variables loading on only one or a few factors is not that important, and may even be undesirable, so varimax rotation may not advantageous.

We repeat again for emphasis that the estimate of $\boldsymbol{\Sigma}_R$ is not changed by rotation. The uniquenesses are also unchanged. Only the loadings change.

Example 18.10. Factor analysis of equity funds: Varimax rotation

The statistical factor analysis in Example 18.9 is repeated here but now with varimax rotation.

```
Call:
factanal(x = equityFunds[, 2:9], factors = 4,
      rotation = "varimax")

Uniquenesses:
EASTEU  LATAM  CHINA  INDIA ENERGY MINING   GOLD  WATER
 0.735  0.368  0.683  0.015  0.005  0.129  0.005  0.778

Loadings:
       Factor1 Factor2 Factor3 Factor4
EASTEU 0.436   0.175   0.148   0.148
LATAM  0.748   0.174           0.180
CHINA  0.494           0.247
INDIA  0.243           0.959
ENERGY 0.327   0.118           0.934
```

```
MINING 0.655   0.637           0.168
GOLD   0.202   0.971
WATER  0.418                   0.188

                Factor1 Factor2 Factor3 Factor4
SS loadings       1.80    1.45    1.03    1.00
Proportion Var    0.23    0.18    0.13    0.12
Cumulative Var    0.23    0.41    0.54    0.66

Test of the hypothesis that 4 factors are sufficient.
The chi square statistic is 17 on 2 degrees of freedom.
The p-value is 2e-04
```

The most notable change compared to the nonrotated loadings is that now all loadings with an absolute value above 0.1 are positive. Therefore, the factors all represent long positions, whereas before some were more like hedge portfolios. However, the rotated factors seem less interpretable compared to the unrotated factors, so a financial analyst might prefer the unrotated factors. □

18.7 Bibliographic Notes

The Fama–French three-factor model was introduced by Fama and French (1993) and discussed further in Fama and French (1995, 1996). Connor (1995) compares the three types of factor models and finds that macroeconomic factor models have less explanatory power than other factor models. Example 18.5 was adopted from Zivot and Wang (2006). Sharpe, Alexander, and Bailey (1999) has a brief description of the BARRA, Inc. factor model. The `yields.txt` data set is from the `Rsafd` package distributed by Professor René Carmona.

18.8 R Lab

18.8.1 PCA

In the first section of this lab, you will do a principal components analysis of daily yield data in the file `yields.txt`. R has functions, which we will use later, that automate PCA, but it is easy to do PCA "from scratch" and it is instructive to do this. First load the data and, to get a feel for what yield curves look like, plot the yield curves on days 1, 101, 201, 301, . . ., 1101. There are 1352 yield curves in the data, so you will see a representative sample of them. The yield curves change slowly, which is why one should look at yield curves that are spaced rather far (100 days) apart.

```
yieldDat = read.table("yields.txt", header = T)
maturity = c((0:5), 5.5, 6.5, 7.5, 8.5, 9.5)
pairs(yieldDat)
par(mfrow = c(4,3))
for (i in 0:11)
{
plot(maturity, yieldDat[100 * i + 1, ], type = "b")
}
```

Next compute the eigenvalues and eigenvectors of the sample covariance matrix, print the results, and plot the eigenvalues as a scree plot.

```
eig = eigen(cov(yieldDat))
eig$values
eig$vectors
par(mfrow = c(1, 1))
barplot(eig$values)
```

The following R code plots the first four eigenvectors.

```
par(mfrow=c(2, 2))
plot(eig$vector[ , 1], ylim = c(-0.7, 0.7), type = "b")
abline(h = 0)
plot(eig$vector[ , 2], ylim = c(-0.7, 0.7), type = "b")
abline(h = 0)
plot(eig$vector[ , 3], ylim = c(-0.7, 0.7), type = "b")
abline(h = 0)
plot(eig$vector[ , 4], ylim = c(-0.7, 0.7), type = "b")
abline(h = 0)
```

Problem 1 *It is generally recommended that PCA be applied to time series that are stationary. Plot the first column of* `yieldDat`*. (You can look at other columns as well. You will see that they are fairly similar.) Does the plot appear stationary? Why or why not? Include your plot with your work.*

Another way to check for stationarity is to run the augmented Dickey–Fuller test. You can do that with the following code:

```
library("tseries")
adf.test(yieldDat[ , 1])
```

Problem 2 *Based on the augmented Dickey–Fuller test, do you think the first column of* `yieldDat` *is stationary? Why or why not?*

Run the following code to compute changes in the yield curves. Notice the use of `[-1,]` to delete the first row and similarly the use of `[-n, ]`.

```
n=dim(yieldDat)[1]
delta_yield = yieldDat[-1, ] - yieldDat[-n, ]
```

Plot the first column of `delta_yield` and run the augmented Dickey–Fuller test to check for stationarity.

Problem 3 *Do you think the first column of* `delta_yield` *is stationary? Why or why not?*

Run the following code to perform a PCA using the function `princomp()`, which is similar, although not identical, to `prcomp()`. By default, `princomp()` does a PCA on the covariance matrix, though there is an option to use the correlation matrix instead. We will use the covariance matrix. The second line of the code will print the names of the components in the object that is returned by `princomp()`. As you can see, the `names` function can be useful for learning just what is being returned. You can also get this information by typing `?princomp`.

```
pca_del = princomp(delta_yield)
names(pca_del)
summary(pca_del)
par(mfrow = c(1, 1)
plot(pca_del)
```

Problem 4 *(a) The output from* `names` *includes the following:*

```
[1] "sdev"  "loadings" "center"  "scores"
```

Describe each of these components in mathematical terms. To answer this part of the question, you can print and plot the components to see what they contain and use `R`*'s help for further information.*

(b) What are the first two eigenvalues of the covariance matrix?

(c) What is the eigenvector corresponding to the largest eigenvalue?

(d) Suppose you wish to "explain" at least 95 % of the variation in the changes in the yield curves. Then how many principal components should you use?

18.8.2 Fitting Factor Models by Time Series Regression

In this section, we will start with the one-factor CAPM model of Chap. 17 and then extend this model to the three-factor Fama–French model. We will use the data set `Stock_Bond_2004_to_2005.csv` on the book's website, which contains stock prices and other financial time series for the years 2004 and 2005. Data on the Fama–French factors are available at Prof. Kenneth French's website

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html#Research

where RF is the risk-free rate and Mkt.RF, SMB, and HML are the Fama–French factors.

Go to Prof. French's website and get the daily values of RF, Mkt.RF, SMB, and HML for the years 2004–2005. It is assumed here that you've put the data in a text file FamaFrenchDaily.txt. Returns on this website are expressed as percentages.

Now fit the CAPM to the four stocks using the lm command. This code fits a linear regression model separately to the four responses. In each case, the independent variable is Mkt.RF.

```
#  Uses daily data 2004-2005

stocks = read.csv("Stock_Bond_2004_to_2005.csv",header=T)
attach(stocks)
stocks_subset = as.data.frame(cbind(GM_AC,  F_AC, UTX_AC, MRK_AC))
stocks_diff = as.data.frame(100 * apply(log(stocks_subset),
   2, diff) - FF_data$RF)
names(stocks_diff) = c("GM", "Ford", "UTX", "Merck")

FF_data = read.table("FamaFrenchDaily.txt", header = TRUE)
FF_data = FF_data[-1, ] # delete first row since stocks_diff
                        # lost a row due to differencing

fit1 = lm(as.matrix(stocks_diff) ~ FF_data$Mkt.RF)
summary(fit1)
```

Problem 5 *The CAPM predicts that all four intercepts will be zero. For each stock, using* $\alpha = 0.025$*, can you accept the null hypothesis that its intercept is zero? Why or why not? Include the p-values with your work.*

Problem 6 *The CAPM also predicts that the four sets of residuals will be uncorrelated. What is the correlation matrix of the residuals? Give a 95 % confidence interval for each of the six correlations. Can you accept the hypothesis that all six correlations are zero?*

Problem 7 *Regardless of your answer to Problem 6, assume for now that the residuals are uncorrelated. Then use the CAPM to estimate the covariance matrix of the excess returns on the four stocks. Compare this estimate with the sample covariance matrix of the excess returns. Do you see any large discrepancies between the two estimates of the covariance matrix?*

Next, you will fit the Fama–French three-factor model. Run the following R code, which is much like the previous code except that the regression model has two additional predictor variables, SMB and HML.

```
fit2 = lm(as.matrix(stocks_diff) ~ FF_data$Mkt.RF +
   FF_data$SMB + FF_data$HML)
summary(fit2)
```

Problem 8 *The CAPM predicts that for each stock, the slope (beta) for* SMB *and* HML *will be zero. Explain why the CAPM makes this prediction. Do you accept this null hypothesis? Why or why not?*

Problem 9 *If the Fama–French model explains all covariances between the returns, then the correlation matrix of the residuals should be diagonal. What is the estimated correlations matrix? Would you accept the hypothesis that the correlations are all zero?*

Problem 10 *Which model, CAPM or Fama–French, has the smaller value of AIC? Which has the smaller value of BIC? What do you conclude from this?*

Problem 11 *What is the covariance matrix of the three Fama–French factors?*

Problem 12 *In this problem, Stocks 1 and 2 are two stocks, not necessarily in the* Stock_FX_Bond_2004_to_2005.csv *data set. Suppose that Stock 1 has betas of 0.5, 0.4, and* -0.1 *with respect to the three factors in the Fama–French model and a residual variance of 23.0. Suppose also that Stock 2 has betas of 0.6, 0.15, and 0.7 with respect to the three factors and a residual variance of 37.0. Regardless of your answer to Problem 9, when doing this problem, assume that the three factors do account for all covariances.*

(a) Use the Fama–French model to estimate the variance of the excess return on Stock 1.
(b) Use the Fama–French model to estimate the variance of the excess return on Stock 2.
(c) Use the Fama–French model to estimate the covariance between the excess returns on Stock 1 and Stock 2.

18.8.3 Statistical Factor Models

This section applies statistical factor analysis to the log returns of 10 stocks in the data set Stock_FX_Bond.csv. The data set contains adjusted closing

(AC) prices of the stocks, as well as daily volumes and other information that we will not use here.

The following R code will read the data, compute the log returns, and fit a two-factor model. Note that `factanal` works with the correlation matrix or, equivalently, with standardized variables.

```
dat = read.csv("Stock_FX_Bond.csv")
stocks_ac = dat[ , c(3, 5, 7, 9, 11, 13, 15, 17)]
n = length(stocks_ac[ , 1])
stocks_returns = log(stocks_ac[-1, ] / stocks_ac[-n, ])
fact = factanal(stocks_returns, factors = 2, rotation = "none")
print(fact)
```

Loadings less than the parameter `cutoff` are not printed. The default value of `cutoff` is 0.1, but you can change it as in "`print(fact,cutoff = 0.01)`" or "`print(fact, cutoff = 0)`".

Problem 13 *What are the factor loadings? What are the variances of the unique risks for Ford and General Motors?*

Problem 14 *Does the likelihood ratio test suggest that two factors are enough? If not, what is the minimum number of factors that seems sufficient?*

The following code will extract the loadings and uniquenesses.

```
loadings = matrix(as.numeric(loadings(fact)), ncol = 2)
unique = as.numeric(fact$unique)
```

Problem 15 *Regardless of your answer to Problem 14, use the two-factor model to estimate the correlation of the log returns for Ford and IBM.*

18.9 Exercises

1. The file `yields2009.csv` on this book's website contains daily Treasury yields for 2009. Perform a principal components analysis on changes in the yields. Describe your findings. How many principal components are needed to capture 98 % of the variability?
2. Perform a statistical factor analysis of the returns in the data set `midcapD.ts` on the book's website. How many factors did you select? Use (18.20) to estimate the covariance matrix of the returns.
3. Verify equation (18.6).
4. Compute the eigenvectors in Example 18.3 and offer an interpretation of the first two eigenvectors.

References

Connor, G. (1995) The three types of factor models: a comparison of their explanatory power. *Financial Analysts Journal*. 42–46.

Fama, E. F., and French, K. R. (1992) The cross-section of expected stock returns. *Journal of Finance*, **47**, 427–465.

Fama, E. F., and French, K. R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**, 3–56.

Fama, E. F., and French, K. R. (1995) Size and book-to-market factors in earnings and returns. *Journal of Finance*, **50**, 131–155.

Fama, E. F., and French, K. R. (1996) Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51**, 55–84.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979) *Multivariate Analysis*, Academic Press, London.

Sharpe, W. F., Alexander, G. J., and Bailey, J. V. (1999) *Investments*, 6th ed., Prentice-Hall, Upper Saddle River, NJ.

Zivot, E., and Wang, J. (2006) *Modeling Financial Time Series with S-PLUS*, 2nd ed., Springer, New York.