8 | Factor Models and Principal Components

The assumption underlying the CAPM is that stock prices move together because of common movement with the market. In Eg. **??**, the empirical evidence indicates that the CAPM beta does not completely explain the expected asset returns. Additional factors may be required to characterized the behavior of expected returns and naturally leads to consideration of multifactor pricing models.

Multifactor Pricing Models

The multifactor pricing model is a linear factor model, the return generating process for asset returns being consider is

$$R_j = a_i + \boldsymbol{\beta}_j \boldsymbol{F} + \varepsilon_j, \qquad E[\varepsilon_j \,|\, \boldsymbol{F}] = 0, \quad \operatorname{var}(\varepsilon_i) = \sigma_j^2. \tag{8.1}$$

where R_j is the return for asset j, a_j is the intercept, $\boldsymbol{\beta}_j$ is a $p \times 1$ vector of factor loadings (or sensitivities) for asset j, $\boldsymbol{F}$ is a $p \times 1$ vector of common factors and ε_j is the disturbance term for asset j. For the system of N assets,

$$\boldsymbol{R} = \boldsymbol{a} + \boldsymbol{B}^T \boldsymbol{F} + \boldsymbol{\varepsilon}, \qquad E[\boldsymbol{\varepsilon} \,|\, \boldsymbol{F}] = \boldsymbol{0}, \qquad \operatorname{var}(\varepsilon) = \boldsymbol{\Sigma}_\varepsilon. \tag{8.2}$$

In the system equation, $\boldsymbol{R} = (R_1 \ldots, R_N)^T$, $\boldsymbol{a}$ is an $N \times 1$ vector, $\boldsymbol{B}$ is a $p \times N$ matrix $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)$.

There are two major theoretical approaches exist. The Intertemporal Capital Asset Pricing Model (ICAPM) developed by Merton (1973) and the Arbitrage Pricing Theory (APT) introduced by Ross (1976).

Intertemporal CAPM. Based on equilibrium argument as the CAPM, Merton (1973) presented a multifactor model having the market portfolio as the main factor and state variables as additional factors. The additional factors arise from investors's demand to hedge uncertainty about future investment opportunities. The class of ICAMP is thus, with $F_1 = R_M$ the market portfolio, and $a_j = (1-\beta_{1j})\mu_f$ in (8.1) and the model can rewritten as

$$R_j = \mu_f + \beta_1(R_M - \mu_f) + \boldsymbol{\beta}_{2j}^T \boldsymbol{F}_2 + \varepsilon_j \ , \tag{8.3}$$

where the $(p-1)\times 1$ vector β_{2j} is $\boldsymbol{\beta}_j$ without the first element and the $(p-1)\times 1$ vector $\boldsymbol{F}_2$ is F without the first common factor. For the system of N assets (8.2) becomes

$$\boldsymbol{R} = \mu_f \mathbf{1}_N + \boldsymbol{\beta}_1(R_M - \mu_f) + \boldsymbol{B}_2^T \boldsymbol{F}_2 + \boldsymbol{\varepsilon} \ . \tag{8.4}$$

where $\mathbf{1}_N$ is a length N vector of 1's, $\boldsymbol{B}_2$ is $\boldsymbol{B}$ without the first column vector. Both (8.3) and (8.4) resemble the CAPM model except the additional second term at the RHS.

Several authors explore this model, notably Fama and French in a series of papers from 1993 to 1998. The Fama-French three factor model adding two fundamental factors to the market portfolio is the most popular among finance practitioners. The historical data

of the fundamental factors in their models are available at Kenneth French's web site at Dartmouth. We will use this model to demonstrate the estimation and tests of ICPM models.

Arbitrage Pricing Theory. The APT was introduced by Ross (1976) as an alternative to the CAPM model. It is more general than the CAPM in that it allows multiple risk factors in the linear factor model. Furthermore, unlike the CAPM, the APT does not require the identification of the market portfolio. It is assumed that the markets are competitive and frictionless and that the expected asset returns are approximated with an unknown number of factors that may or may not be observable. The APT is derived from a statistical model for the returns given in (8.1) and (8.2) with much less stringent assumptions than the CAPM with the main assumption being no arbitrage. And under the absence of arbitrage, it is shown that

$$\boldsymbol{\mu} = \lambda_0 \mathbf{1}_N + \boldsymbol{B}^T \boldsymbol{\lambda}_p \tag{8.5}$$

where $\boldsymbol{\mu}$ is the $N\times 1$ expected return vector, λ_0 is the zero-beta parameter and is the riskfree return if such an asset exists, and $\boldsymbol{\lambda}_p$ is a$p\times 1$ vector of factor risk premia. There are $N-p-1$ restrictions in (8.5), these are the main restrictions of the APT. The APT factors can be defined in many ways including the ICAMP.

Another key assumption on the multifactor models is that all the included factors are pervasive. That is, all the factors play and important role in explaining the returns of the of the assets, otherwise, risk would be reduced and return would be aways be equal to the zero-beta or risk-free rate.

Model for Estimation and Testing

As in the CAPM, we define a model for a times series of length n of returns on N assets that obeys the same linear factor model.

$$Y_{jt} = \alpha_j + \boldsymbol{\beta}_j^T \boldsymbol{F}_t + \varepsilon_{jt}, \qquad j = 1, \ldots, N \text{ and } t = 1, \ldots, n. \quad (8.6)$$

where Y_{jt} is normally the excess return of asset j, but Y_{jt} would occasionally be referred as the return since the framework considered here is more general. In matrix form,

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \boldsymbol{B}^T \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \qquad E[\boldsymbol{\varepsilon}_t \mid \boldsymbol{F}_t] = \boldsymbol{0}, \qquad E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T \mid \boldsymbol{F}_t] = \boldsymbol{\Sigma}_\varepsilon \quad (8.7)$$

A strong assumption on $\boldsymbol{\Sigma}_\varepsilon$ is that it is diagonal, this is not likely to be satisfied when N is large. There are several different types and approaches of factor models.

1. Observable factors – The factors $\boldsymbol{F}$ are returns to observed traded portfolios on the basis of a security characteristic such as size and value. Or the factor $\boldsymbol{F}$ are observed macro variables such as yield spread.
2. Observable characteristics – Models in which the $\boldsymbol{\beta}_j$ are observable characteristics of depend in a simple way on observable characteristic such as industry or country.
3. Statistical factor models – Models in which both $\boldsymbol{F}$ and $\boldsymbol{\beta}_j$ are unknown quantities.

Cases 1 and 2, the model is essentially a linear regression model. Case 1 can be viewed a generalization of the CAPM. Case 3 is quite different what we have discussed, it is no longer a regression model.

Models with Observable Factors

There are several macroeconomic factor models built on the work of Chen, Roll and Ross (1986). The factors used consist of unexpected changes or surprises of macroeconomic variables. Here unexpected changes denote the residuals of the macroeconomic variables after removing their dynamic dependence.

For the case of observable factors, we will focus on the Fama-French models. Fama and French found that both size and the ratio of book value of equity to the market value of equity have a strong role in determining the cross section of average return on common stocks. Adding these two factors to the CAPM, they proposed the three factor models.

Three-Factor Model. For a given asset fundamental (e.g., ratio of book-to-market value), Fama and French (1992) determined factor realizations using a two-step procedure. First, they sorted the assets based on the observed values. Then they formed a hedge portfolio which is long in the top quintile (1/3) of the sorted assets and short in the bottom quintile of the sorted assets. The observed return on this hedge portfolio at time t is the observed factor realization for the given asset fundamental. The procedure is repeated for each asset fundamental under consideration. The three fundamentals used by Fama and French are:

1. The excess returns of the market portfolio, which are the value-weighted returns on all NYSE, AMEX and NASDAQ stocks minus the 1-month T-Bill rates.

2. SMB (Small Minus Big) measures the "size effect". It is the difference of returns between large and small capitalization.
3. HML (High Minus Low) measures the "value effect". It is the difference of returns between high and low book-to-market ratios.

Both SMB and HML are based on two fixed portfolios (small and big, high and low). The factors depend only on the financial markets and the economy as a whole.

Five-Factor Model. Fama and French (2015) extended their model by adding two additional factors, profitability and investment factors.

4. RMW (Robust Minus Weak) measures the "profitability effect". It is the difference between the returns of the most profitable (robust) firms and the least profitable (weak) firms.
5. CMA (Conservative minus Aggressive) measures "investment effect". It is the difference between the returns of firms that invest conservatively and firms that invest aggressively.

The study done by Fama and French (2015) found that their five factor model that the value effect factor HML is redundant for describing average returns when profitability and investment factors have been added into the equation due to the correlations among these factor variables. In some applications, a four-factor model may be used instead of the five-factor model.

Estimation and Testing For the French-Fama models, (8.6) and (8.7) Y_{jt} and $\boldsymbol{Y}_t$ are the models of excess returns, that is,

$$Y_{jt} = R_{jt} - \mu_f \qquad \text{and} \qquad \boldsymbol{Y}_t = \boldsymbol{R}_t - \mu_f \mathbf{1}_N$$

respectively, where μ_f is return of the risk-free asset. Therefore,

$$\boldsymbol{\Sigma}_R = \text{var}(\boldsymbol{R}_t) = \boldsymbol{B}^T \boldsymbol{\Sigma}_F \boldsymbol{B} + \boldsymbol{\Sigma}_\varepsilon \tag{8.8}$$

Imposing the strong assumption that $\boldsymbol{\Sigma}_\varepsilon = \text{diag}\{\sigma^2_{\varepsilon,1}, \ldots \sigma^2_{\varepsilon,N}\}$, the variance of the jth asset return is

$$\text{var}(R_{jt}) = \boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_F \boldsymbol{\beta}_j + \sigma^2_{\varepsilon,j} \tag{8.9}$$

and the covariance between the the ith and jth asset returns is

$$\text{Cov}(R_{it}, R_{jt}) = \boldsymbol{\beta}_i^T \boldsymbol{\Sigma}_F \boldsymbol{\beta}_j\,. \tag{8.10}$$

The quantity $\boldsymbol{\beta}_j^T \boldsymbol{\Sigma}_F \boldsymbol{\beta}_j$, which is the portion of the variance of R_{jt} contributed by the p common factors, is called *the communality*. The remaining portion $\sigma_{\varepsilon,j}$ of the variance of R_{jt} is called *the uniqueness* or *specific variance*.

The CAPM is a factor model where $p = 1$ and F_{1t} is the excess return on the market portfolio. 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, e.g. macro economic factors of Chen, Roll and Ross Models

With sufficient factors, most, perhaps all commonalities between assets should be accounted for in the model. Then (1) the intercept

α_j should be zero; and (2) the ε_{jt} should represent factors truly unique to the individual assets and therefore should be uncorrelated across j assets as is being assumed in (8.8).

In order to use the factor model (8.7), one needs estimates of all the parameters in the model, $\boldsymbol{\alpha}$, $\boldsymbol{B}$, $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_\varepsilon$.

Fitting Factor Models by Regression

Suppose $\boldsymbol{F}_t$ are known and observable such as the CAPM, each excess asset return Y_{jt} is a regression model , Y_{jt} on $F_{1t},\ldots,F_{pt}$. Combing all N regressions, (8.7) is simply the multivariate regression model. Then R's `lm()` function can be used to fit a set of returns to factors. The MLE of coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{B}$ are the same as the least squares estimates from the regression. The estimate of $\boldsymbol{\Sigma}_\varepsilon$, the variance-covariance of $\boldsymbol{\varepsilon}_t$ is

$$\hat{\boldsymbol{\Sigma}}_\varepsilon = \frac{1}{n}\sum_{t=1}^{n}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t^T and \quad \hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{Y}_t - \hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{B}}^T\boldsymbol{F}_t\,. \tag{8.11}$$

where $\hat{\boldsymbol{\alpha}} = \overline{\boldsymbol{Y}} - \hat{\boldsymbol{B}}^T\overline{\boldsymbol{F}}$,

$$\hat{\boldsymbol{B}} = \left\{\sum_{t=1}^{n}(\boldsymbol{Y}_t-\overline{\boldsymbol{Y}})(\boldsymbol{F}_t-\overline{\boldsymbol{F}})^T\right\}^T\left\{\sum_{t=1}^{n}(\boldsymbol{Y}_t-\overline{\boldsymbol{Y}})(\boldsymbol{Y}_t-\overline{\boldsymbol{Y}})^T\right\}^{-1}.$$

For each regression of asset j, the OLS estimator $(\hat{\alpha},\hat{\boldsymbol{\beta}}_j^T)^T$ is normally distributed with mean $(\alpha,\boldsymbol{\beta}_j^T)^T$ and variance

$$\mathrm{var}[(\hat{\alpha},\hat{\boldsymbol{\beta}}_j^T)^T] = \boldsymbol{M}\sigma^2_{j\varepsilon}, \qquad \boldsymbol{M} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}, \qquad \boldsymbol{X} = [\boldsymbol{1}_n, \boldsymbol{F}]$$

where $\boldsymbol{F}$ is an $n\times p$ matrix each row is $\boldsymbol{F}_t$, $t=1,\ldots,n$. Thus, the distribution of $\hat{\boldsymbol{\alpha}}$ is normal and,

$$E[\hat{\boldsymbol{\alpha}}] = \boldsymbol{\alpha} \qquad \text{and} \qquad \mathrm{var}(\hat{\boldsymbol{\alpha}}) = m_{11}\boldsymbol{\Sigma}_\varepsilon\,. \tag{8.12}$$

where m_{11} be the (1,1)th entry of $\boldsymbol{M}$.

Estimating Covariances of Asset Returns The variance-covariance matrix of $\boldsymbol{R}_t$ based on the factor model 8.7 is given by 8.8 and can be estimated by

$$\hat{\boldsymbol{\Sigma}}_R = \hat{\boldsymbol{B}}^T\tilde{\boldsymbol{\Sigma}}_F\hat{\boldsymbol{B}} + \mathrm{diag}\{\hat{\boldsymbol{\Sigma}}_\varepsilon\}, \tag{8.13}$$

where $\tilde{\boldsymbol{\Sigma}_F}$ is sample variance and $\hat{\boldsymbol{\Sigma}}_\varepsilon$ is the MLE of (8.11).

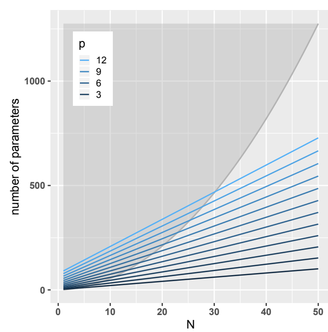
For smaller dimensional asset returns N, the sample covariance is preferable than the factor model covariance (8.13). The advantage of using a factor model to estimate the covariance of asset returns is when a portfolio consists of a large number of assets, N. The sample covariance is unbiased, however, it contains $N(N+1)/2$ estimates, each with error. When N is large, the result can be a sizable loss of precision.

In contrast, the factor model requires estimates of $N\times p$ parameters in $\boldsymbol{B}$, $p(p+1)/2$ parameters in $\boldsymbol{\Sigma}_F$ and N parameters in the diagonal matrix $\boldsymbol{\Sigma}_\varepsilon$ for a total of $(p+1)(N+p/2)$. Typically, N, the number of assets is a large number and p, the number of common factors is much smaller, the reduction in parameter numbers can be substantial.

The downside of the factor model is that there will be bias in the estimate of if the model is misspecified, especially if $\boldsymbol{\Sigma}_\varepsilon$ is not diagonal. This is an example of bias-variance tradeoff.

```
## === ===> Number of Parameters <======
##
##  Sample Variance:
##       N = 3 N = 6 N = 9 N = 15 N = 27 N = 51 N = 99
## S2        6     21     45     120     378    1326    4950
## Factor Models:
##       N = 3 N = 6 N = 9 N = 15 N = 27 N = 51 N = 99
## p = 3     18     30     42      66     114     210     402
## p = 5     33     51     69     105     177     321     609
```

For $p = 3$, N needs to be at least 9 to have smaller number of parameter in the factor model. The plot shown here are number of parameters for $N = 1,\ldots,30$ and $p = 1,\ldots,12$. The curve is the number of parameters of sample covariance.



Model validation. Just like the CAPM, an ICAPM such as the Fama-French models assume that no asset is systematically under or over priced, implying that the intercept $\boldsymbol{\alpha}$ in the model (8.7) is zero. Given that the expected errors are zero, the expected return of an asset is completely determined by the market factors.

For example, in the F-F 3 factor model, the expected return of asset $jE(R_{jt}) = \mu_f + \beta_{1j}(\mu_M - \mu_f) + \beta_{2j}E(\mathrm{SML}_t) + \beta_{3j}E(\mathrm{HML}_t)$ if the 3-factor pricing model holds. Model validation is simply testing H_0 :

$\alpha_i = 0$, this test for an individual asset is available form any standard regression output.

The null hypothesis of testing the factor pricing model holds for all the N asset is

$$H_0 : \boldsymbol{\alpha} = \mathbf{0} \tag{8.14}$$

The Wald and likelihood ratio test statistics are analogous to those for the CAPM with adjustments for the number of factors p.

Wald statistic Following from (8.12) and Appendix of Handout 7, the Wald statistic is

$$\mathscr{T}_W = \frac{n-N-p}{nN}\,\frac{1}{m_{11}}\,\hat{\boldsymbol{\alpha}}^T\hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}\hat{\boldsymbol{\alpha}} \sim F_{N,n-N-p}\,, \tag{8.15}$$

where $\hat{\boldsymbol{\alpha}}$ is the least squares estimate and $\hat{\boldsymbol{\Sigma}}_\varepsilon$ is (8.11)

Likelihood ratio statistic The restricted model under the null (8.14) is the regression model (8.7) with $\boldsymbol{\alpha} = \mathbf{0}$. Let $\hat{\boldsymbol{\varepsilon}}_{0t}$ be the residuals from regressing $\boldsymbol{Y}_t$ on $\boldsymbol{F}_t$ without intercept and $\hat{\boldsymbol{\Sigma}}_{\varepsilon,0} = n^{-1}\sum_t \boldsymbol{\varepsilon}_{0t}\boldsymbol{\varepsilon}_{0t}^T$. The likelihood ratio test statistic is

$$\mathscr{T}_{\mathrm{LR}} = \big(n - N/2 - p - 1\big)\big\{\log|\hat{\boldsymbol{\Sigma}}_{\varepsilon,0}| - \log|\hat{\boldsymbol{\Sigma}}_\varepsilon|\big\} \sim \chi^2_N\,. \tag{8.16}$$

Eg 8.1. Fitting the Fama-French model to daily returns of 18 assets, from January 1, 2015 to August 30, 2024, $n = 2431$. The 18 stocks include 5 from Tech sector – Apple, Microsoft, Cisco, IBM, Intel; 4 from Oil & Gas Industry – Chevron, Exxon Mobil, Enterprise Products, Phillips 66; 4 from banking industry, Citi, Bank of America, JP Morgan, Wells Fargo; and 5 from pharmaceutical industry – Merck, Abbott, Bristol-Myers Squibb, Eli Lilly, Pfizer, Johnson& Johnson. The daily Fama-French factors are taken from Professor Kenneth French's website.

```
cat("Starting date:"); head(Rt[,1:5],2);cat("\nEnding date:"); tail(Rt[,1:5],2)

## Starting date:
##                 AAPL     MSFT      CSCO     IBM    INTC
## 2015-01-05 -2.8575943 -0.92381 -2.012180 -1.5860 -1.1340
## 2015-01-06  0.0094206 -1.47862 -0.036963 -2.1802 -1.8813
##
## Ending date:
##               AAPL    MSFT    CSCO     IBM   INTC
## 2024-08-29  1.44650 0.61186 1.07742 0.22146 2.6172
## 2024-08-30 -0.34438 0.96838 0.29724 1.61089 9.0648

cat("dimension of Rt:", dim(Rt))

## dimension of Rt: 2431 18
```

The object `Rt` is daily returns in percentage (%). All the return data in this handout are in percentage as the Fama French factor data. The 3 factor data has 4 columns, `Mkt.RF` ($R_M - \mu_f$), `SMB`, `HML`, `RF` (μ_f) of daily data from July 1, 1926 to Aug 31, 2024.

```
FF_data = read.table("F-F_Research_Data_Factors_daily.txt", header=T)

ind.0 = which(rownames(FF_data) ==20130103)
FF_data = FF_data[-(1:(ind.0-1)),]
cat("dimension of FF_data:", dim(FF_data))

##          Mkt.RF  SMB   HML  RF
## 20150105  -1.84 0.33 -0.68   0
##          Mkt.RF   SMB  HML    RF
## 20240830   0.98 -0.55 0.04 0.022

cat("dimension of FF_data:", dim(FF_data))

## dimension of FF_data: 2431 4
```


The object `FF_data` now has the same time period as that of `Rt`. Before fitting the data to the regression model (8.7), the risk free rate series `RF` is subtracted from `Rt` to yield excess returns `Yt`. It is convenient to "attach" `FF_data`, so we can retrieve each column simply with its column name.

```
attach(FF_data)
Yt = apply(Rt,2, function(x) x-RF); dimnames(Yt)[[2]] = syb;
n = dim(Yt)[1]; N = dim(Yt)[2]; p = 3
fit = lm(Yt~Mkt.RF+SMB+HML); sfit = summary(fit)
```

Just like the CAPM fit, the output `fit` contains 18 Fama French 3 factor model fit for each return series. Figure 8.1 shows the R Squared values and betas of the three factors for all 18 assets. Of 18 assets, 11 have excess-market-return betas that exceed 1, including all 4 banks and 4 of 5 tech companies. The 4 bank assets also have high R-Squared values ranging from 0.6784 to 0.7488.

```
industry = c("Tech", "O&G", "Bank", "Drug")
Ns = c(5,4,4,5); by_industry = c();
for(i in 1:4) by_industry = c(by_industry, rep(industry[i],Ns[i]))
table(Aggressive = coef(fit)[2,] > 1, by_industry)

##           by_industry
## Aggressive Bank Drug O&G Tech
##      FALSE    0    5   2    2
##      TRUE     4    0   2    3

R.Squared = c(); for(i in 1:N) R.Squared[i] = sfit[[i]]$r.squared
table(Hi_R.Sq = R.Squared > 0.5, by_industry)

##        by_industry
## Hi_R.Sq Bank Drug O&G Tech
##   FALSE    0    5   2    3
##   TRUE     4    0   2    2
```

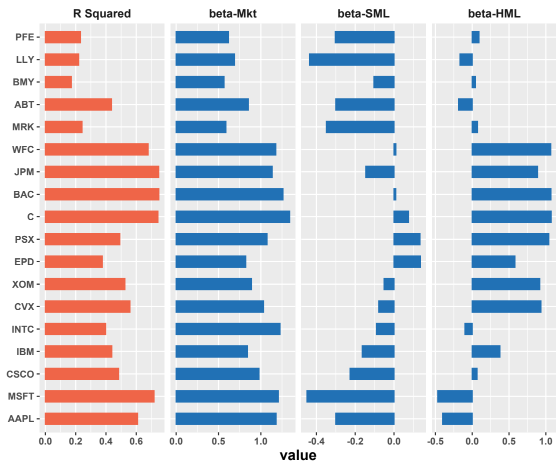


Figure 8.1: R-Squared and betas of excess daily returns of 18 assets by fitting Fama-French three factor model.

Most of the SMB betas are negative due to the fact that this factor measures "small firm effect" and our daily returns are from large firms. Negative HML betas are an indication that the assets are growth stocks, which have potential for growth in the foreseeable future, Microsoft and Apple have the lowest HML betas.

The null hypothesis of validating the factor pricing model is zero-intercept. For individual asset, only three, Intel, Citi and Eli Lilly, out of 18 assets are found significant at 5% in testing $H_0 : \alpha_j = 0$. The factor pricing model does not hold for the Intel, Citi and Eli Lilly daily returns.

```
Alpha = c()
for(i in 1:N){
    Alpha= rbind(Alpha, sfit[[i]]$coef[1,])
}
dimnames(Alpha)[[1]] = syb  ## syb is the vector of tickers
```



```
head(Alpha,1)

##       Estimate Std. Error t value Pr(>|t|)
## AAPL 0.024428   0.023014  1.0614   0.2886

cat("alpha significant at 5%:", syb[(alpha[,4] < 0.05)]);
 alpha[(alpha[,4]< 0.05),]

## alpha significant at 5%: INTC C LLY
##         Estimate Std. Error t value Pr(>|t|)
## INTC -0.075957   0.035508 -2.1391 0.032523
## C    -0.045104   0.021413 -2.1064 0.035274
## LLY   0.072218   0.031193  2.3152 0.020685
```

To test all 18 asset returns follow the factor pricing model, we test $H_0 : \boldsymbol{\alpha} = 0$ with both the Wald and likelihood ratio tests. As expected, both tests are not significant with very similar p-values. This collection of 18 assets follows the 3-factor pricing model. Conditioning on the factors $\boldsymbol{F}_t$, their expected return is simply $\boldsymbol{\mu}_f + \boldsymbol{B}^T \mu_F$.

```
#**** Wald Test for all alpha's are zero ****
alpha= Alpha[,1]    ## the vector of alpha-hat
res = resid(fit); Sig.e = 1/n*t(res)%*%res   ## Sigma_epsilon hat
m11 = sfit[[1]]$cov.unscaled[1,1]
wald = (n-N-p)/N*1/(m11*n)*(t(alpha)%*%solve(Sig.e)%*%alpha)  ## Wald statistic
c(statistic = wald, p.value = 1-pf(wald, N, n-N-p))

## statistic   p.value
##   1.62685   0.04595
```

The Wald test has a p-value. The expected returns of the 18 assets follows the Fama-French 3-factor model. Similar result from the likelihood ratio test. The p-values of the two tests are almost identical.

```
#***** Likelihood Ratio Test for all alpha's are zero ****
res.0 = resid(lm(Yt~Mkt.RF+SMB+HML-1))   ## residuals from the restricted model
Sig.e0 = 1/n*t(res.0)%*%res.0  ## Sigma_epsilon hat under H0
lr = (n-N/2-p-1)*(log(det(Sig.e0))-log(det(Sig.e))) ## LRT statistic
c(statistic = lr, p.value =1-pchisq(lr, N))

## statistic   p.value
## 29.203458  0.045949
```

Tests for a subset of returns can be easily done with a minor modification from the last two tests. For example, the three assets that are significant in testing $H_0 : \alpha_j = 0$ are from three different industries, we may like to have model validation test for each industry. Since the tests are used repeatedly, putting these R code can be useful. The R code of R functions used here are posted on Canvas.

```
args(wald.fun)

## function (est, est.var, n, p, W = diag(length(est)))

args(lrt.fun)

## function (sig, sig0, n)

cat("testing alphas are zero for each industry")

## testing alphas are zero for each industry

wald = c(); lrt = c()
for(i in industry){
  ind = which(by_industry == i)
  wald = rbind(wald,  wald.fun(alpha[ind], m11*Sig.e[ind,ind], n = n, p = p) )
  lrt = rbind(lrt,      lrt.fun(Sig.e[ind,ind], Sig.e0[ind,ind], n = n))
}
rownames(wald) = rownames(lrt) = industry
cat("Wald test by industry:"); wald

## Wald test by industry:
##         Wald  p.value df1  df2
## Tech 1.70179 0.130781   5 2423
## O&G  0.25261 0.908179   4 2424
## Bank 2.54413 0.037826   4 2424
## Drug 1.87725 0.095047   5 2423
```

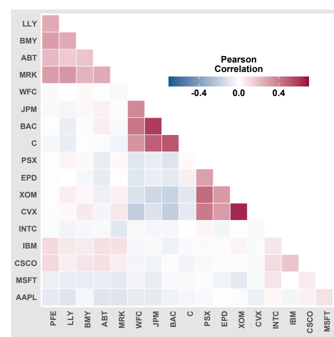


```
cat("LRT by industry:"); lrt

## LRT by industry:
##          LRT  p.value df
## Tech  8.5063 0.130452  5
## O&G   1.0115 0.908052  4
## Bank 10.1678 0.037694  4
## Drug  9.3817 0.094776  5
```

Now we turn to the covariance estimation. The is the example that the number of parameters is substantial reduced with the model based estimation. The data has $N = 18$ asset returns and $p = 3$ factors. The sample covariance has 171 parameters, while the 3 factor model covariance has only 78 parameters. This reduction, however, relies on the assumption that the components of the errors $\boldsymbol{\varepsilon}_t$ are not correlated.

```
## covariance of asset returns based on  the Factor model
Bhat = coef(fit)[-1,] ## betas
Sig.R = t(Bhat)%*%F.S2%*%Bhat + diag(diag(Sig.e));
```

The sample correlation of residuals are between -0.1952 and 0.6422, many are close to zero as shown in the heatmap here. The pair-wise correlation tests shown that 76 out of 153 pairs are significant at 1% and 88 are at 5%. We stress that these tests are not the same as testing if covariance matrix of



$\boldsymbol{\varepsilon}_t$ is diagonal. The test for a covariance matrix being diagonal or structured can be done by the likelihood ratio tests.

Tests for block-diagonal matrices. Suppose that the d-dimensional random vector $\boldsymbol{y} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The hypothesis that the subvectors $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ with lengths $d_1, \ldots, d_k$ are independent implies $\boldsymbol{\Sigma}_{ij} = 0$, for $i \neq j$, that is,

$$H_0 : \boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\Sigma}_{11}, \ldots, \Sigma_{kk}\}$$

The LRT statistic is

$$\text{LRT} = -\log \frac{\det(\hat{\boldsymbol{\Sigma}})}{\det(\hat{\boldsymbol{\Sigma}_{11}}) \cdots \det(\hat{\boldsymbol{\Sigma}_{kk}})}.$$

For large sample size, with an adjustment factor c the resulting cLRT is approximately χ^2_ν distribution, the degrees of freedom

$$\nu = \frac{1}{2}\left(d^2 - \sum_{i=1}^{k} d_i^2\right).$$

The test for a diagonal matrix is the same test with $d_i = 1$, $i = 1, \ldots, d$. The LRT statistic is simply $-\log\det(\widehat{\text{Corr}}(\boldsymbol{y}))$ and the degrees of freedom $\nu = \frac{1}{2}d(d-1)$.

Eg 8.2. Testing $\hat{\boldsymbol{\Sigma}}_\varepsilon$. We will test for $H_0 : \boldsymbol{\Sigma}_\varepsilon = \text{diag}\{\sigma_{\varepsilon,11}, \ldots, \sigma_{\varepsilon,NN}\}$, and $H_0 : \boldsymbol{\Sigma}_\varepsilon = \text{diag}\{\boldsymbol{\Sigma}_{\varepsilon,11}, \ldots, \boldsymbol{\Sigma}_{\varepsilon,kk}\}$, The R code is posted on Canvas

```
args(cov.diag.test)
## function (Sig, Ns, n, p)
cov.diag.test(Sig.e, Ns= rep(1,N), n = n, p = p)
## *** Testing if the matrix is diagonal ***
## LRT -statistic: 6819.4   p-value: 0  DF: 153
cov.diag.test(Sig.e, Ns= Ns, n = n, p = p)
## *** Testing if the matrix is block diagonal ***
## LRT -statistic: 668.17   p-value: 0  DF: 121
```


Both tests are significant. However, we can see from the plot that there exists block structure in the variance matrix. The positive correlations in the diagonal blocks reflects the correlation between two returns from the same industry is higher than those from different industries. This suggests that the data can be modeled by the industry factor model, which is an example of model with observable characteristic factors.

Models with Observable Characteristics

Rosenberg (1974) introduced a multi-factor model where $\boldsymbol{F}$ in (8.1) and (8.2) is treated unknown and $\boldsymbol{B}$ is observable and time invariant. The model is a multiple linear regression with N observations and p unknowns, thus the model requires N to be large. For each t, the model can be written as a cross-sectional regression where the $p \times 1$ will be estimated,

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \boldsymbol{B}^T \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t, \qquad \text{var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}_\varepsilon = \text{diag}\{\sigma^2_{11}\}, \ldots, \sigma^2_{NN}\}.$$

OLS Estimation. The OLS estimator of $\boldsymbol{F}_t$ is

$$\hat{\boldsymbol{F}}_{t,\text{OLS}} = (\boldsymbol{B}\boldsymbol{B}^T)^{-1}\boldsymbol{B}\boldsymbol{Y}_t, \quad \hat{\boldsymbol{\alpha}}_{\text{OLS}} = (\mathbf{I}_N - \boldsymbol{B}^T(\boldsymbol{B}\boldsymbol{B}^T)^{-1}\boldsymbol{B})\overline{\boldsymbol{Y}} \qquad (8.17)$$

The estimator of $\hat{\boldsymbol{\Sigma}}_{\varepsilon,\text{OLS}}$ is the diagonal elements of the sample variance matrix based on the residuals,

$$\tilde{\boldsymbol{\varepsilon}}_{t,\text{OLS}} = \boldsymbol{Y}_t - \boldsymbol{B}^T \hat{\boldsymbol{F}}_{t,\text{OLS}}, \quad t = 1, \ldots, n.$$

The sample variance computation is the same with or without $\hat{\boldsymbol{\alpha}}$. The variance matrix of $\boldsymbol{F}_t$ is the sample variance based on $\hat{\boldsymbol{F}}_{t,\text{OLS}}$,

$t=1,\ldots,n$. The estimator of $\boldsymbol{\Sigma}_R$, the variance of R_t, is

$$\hat{\boldsymbol{\Sigma}}_{R,\text{OLS}} = \boldsymbol{B}^T\hat{\boldsymbol{F}}_{t,\text{OLS}}\boldsymbol{B} + \text{diag}\{\hat{\boldsymbol{\Sigma}}_{\varepsilon,\text{OLS}}\}$$

GLS Estimation. The GLS (Generalized LS) estimator is given by

$$\hat{\boldsymbol{F}}_{t,\text{GLS}} = (\boldsymbol{B}\text{diag}\{\hat{\boldsymbol{\Sigma}}_{\varepsilon,\text{OLS}}\}^{-1}\boldsymbol{B}^T)^{-1}\boldsymbol{B}\text{diag}\{\hat{\boldsymbol{\Sigma}}_{\varepsilon,\text{OLS}}\}^{-1}\boldsymbol{Y}_t\,.$$

Similarly, the GLS estimator of $\boldsymbol{\Sigma}_\varepsilon$ is based on the residuals,

$$\tilde{\boldsymbol{\varepsilon}}_{t,\text{GLS}} = \boldsymbol{Y}_t - \boldsymbol{B}^T\hat{\boldsymbol{F}}_{t,\text{GLS}} \qquad \boldsymbol{t}=1,\ldots,n\,.$$

The GLS version of the estimator of $\boldsymbol{\Sigma}_R$, the variance of R_t, is

$$\hat{\boldsymbol{\Sigma}}_{R,\text{GLS}} = \boldsymbol{B}^T\hat{\Sigma}_{F,\text{GLS}}\boldsymbol{B} + \text{diag}\{\hat{\boldsymbol{\Sigma}}_{\varepsilon,\text{GLS}}\}$$

where $\hat{\Sigma}_F$ is the sample variance matrix based on $\hat{\boldsymbol{F}}_{t,\text{GLS}}$, $t=1,\ldots,t$.

Rosenberg founded Barra Inc. which has become one of the leading commercial providers of factor models. This class of models are sometimes referred to as the Barra Factor Model.

Eg 8.3. Industry Factor Model with the data of Eg. 8.1. The $N=18$ assets are classified into 4 industries, tech, oil & gas, banking and pharmaceutical. The $\boldsymbol{B}$ matrix are indicators for the 4 industries,

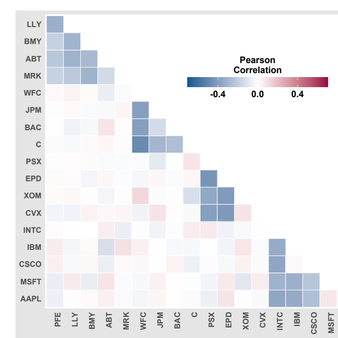
$$Y_{jt} = \alpha_j + \boldsymbol{\beta}_j^T\boldsymbol{F}_t + \varepsilon_{jt}, \qquad j=1,\ldots,N.$$

where $\boldsymbol{\beta}_j = (\beta_{1j},\beta_{2j},\beta_{3j},\beta_{4j})^T$ with $\beta_{ji}=1$ if asset j is in the ith industry and 0 otherwise. For example, if j = Citi, then $\boldsymbol{\beta}_j = (0,0,1,0)^T$.


```
## Making the B matrix
B = outer(industry, by_industry, "==")*1
rownames(B) = industry; colnames(B) = syb; B

## B-Transpose:
##      AAPL MSFT CSCO IBM INTC CVX XOM EPD PSX  C  BAC JPM WFC MRK ABT BMY LLY PFE
## Tech    1    1    1   1    1   0   0   0   0  0    0   0   0   0   0   0   0   0
## O&G     0    0    0   0    0   1   1   1   1  0    0   0   0   0   0   0   0   0
## Bank    0    0    0   0    0   0   0   0   0  1    1   1   1   0   0   0   0   0
## Drug    0    0    0   0    0   0   0   0   0  0    0   0   0   1   1   1   1   1
```

We will only show the OLS estimation, the GLS version has only minimal improvement.

```
#****   Industry Factor Model   ****
Ft = solve(B%*%t(B))%*%B%*%t(Yt) ## dimension 4 x n
Ft = t(Ft) ## dimension n x 4
et = Yt - Ft%*%B  ## dimension n x N
Sig.e = diag(diag(var(et)));  ## Sig.e is diagonal
resid.summary(et, plot = F) ## checking off diagonal entries

##
## Significant pairs at 1% level:  58 of  153 pairs
## Significant pairs at 5% level:  73 of  153 pairs

Sig.R = t(B)%*%var(Ft)%*%B + Sig.e
```



The closer the sample covariance of the residuals $\hat{\boldsymbol{\Sigma}}_\varepsilon$ to a diagonal matrix, the better the estimator $\hat{\boldsymbol{\Sigma}}_R$ is for $\boldsymbol{\Sigma}_R$. There are 58 significant pairs at 1% comparing to 76 of the Fama-French three factor model. This is an indication that the industry factor model gives better estimator for $\boldsymbol{\Sigma}_R$.

Note that the residual variance matrix is singular in this model, the diagonal test does not apply.

Statistical Factor Models

There are two main approaches to choose factors. One is based on the financial economic theory as we have illustrated. The other is statistical approach which builds factors from a comprehensive set of asset returns via *factor analysis* and *principal component analysis*.

Principal components analysis and factor analysis are classical statistical tools. The two topics are different when the number of variables is small, but are approximately the same when the number of variables are large.

In a statistical factor model, neither the factor values $\boldsymbol{F}_t$ nor the loadings $\boldsymbol{B}$ need to be directly observable.

Principal Component Analysis

The PCA can be directly used to the returns $\boldsymbol{R}_t$ to construct a factor model or to reduce the number of risk factors for observable factors. Suppose that there are a large collection of potential risk factors can be included in a factor model, we can use the principal component analysis to blend all risk factors into fewer risk factors.

Blending variables As mention earlier, in the study of their five factor model, Fama and French (2015) found that the value effect factor `HML` becomes redundant when two new factors `RMW` and `CMA` are added. The excess return of the Market portfolio is apparently the major factor, instead of dropping `HML`, we will blend the rest of 4 factors into 3 factors in the next example with data of Eg. 8.1.

Eg 8.4. The 5 factors data are from Kenneth French's website. The data have been trimmed to the same time period as the 18 asset return data.

```
cat("Fama-French 5 factors:"); head(FF5_data,2); tail(FF5_data,1)

## Fama-French 5 factors:
##          Mkt.RF   SMB   HML  RMW   CMA RF
## 20150105  -1.84  0.25 -0.68 0.16 -0.08  0
## 20150106  -1.04 -0.78 -0.31 0.53  0.02  0
##          Mkt.RF   SMB  HML  RMW   CMA    RF
## 20240830   0.98 -0.56 0.04 -0.2 -0.14 0.022
```

Before we apply PCA to the 4 factors `SMB`, `HML`, `RMW` and `CMA`, we first remove the correlation of the market portfolio excess returns from them by regressing them on the factor `Mkt.RF`. The residual vectors of these regressions are uncorrelated to `Mkt.RF`.

```
attach(FF5_data)
Zt = lm(cbind(SMB,HML,RMW,CMA)~Mkt.RF)$resid
pca = prcomp(Zt); summary(pca)

## Importance of components:
##                          PC1   PC2   PC3    PC4
## Standard deviation     0.992 0.687 0.410 0.3319
## Proportion of Variance 0.568 0.272 0.097 0.0635
## Cumulative Proportion  0.568 0.840 0.936 1.0000

head(pca$x,2)

##       PC1      PC2     PC3       PC4
## 1 0.63801  0.54944 0.41768 -0.071448
## 2 0.46026 -0.69892 0.19643  0.011195
```

The four factors, `Mkt.RF`, `PC1`, `PC2` and `PC3`, are uncorrelated each other. We now fit the regression models with these factors to the

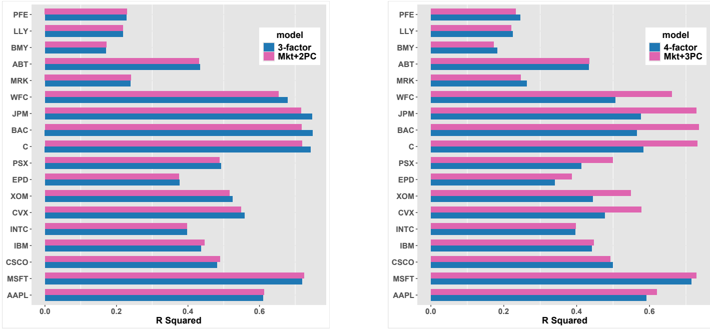


Figure 8.2: R-Squared values. The left plot is the Fama French 3 factor model and the excess market portfolio return + 2 principal components. The right plot is Fama French 4 factor model and the excess market portfolio return + 3 principal components.

daily return data. We will also fit a model with `Mkt.RF` and the first 2 principal components `PC1` and `PC2` to compare with the Fama-French 3 factor model.

```
attach(data.frame(pca$x))
fit.pc2 = lm(Yt~Mkt.RF+PC1+PC2)      # Mkt portfolio returns and 2 PCs
fit.pc3 = lm(Yt~Mkt.RF+PC1+PC2+PC3) # Mkt portfolio returns and 3 PCs
```

Figure 8.2 shows R Squared values of the models. The left plot shows the `Mkt.RF` + 2PC model and the Fama French 3 factor model, the R Squared values of both models are comparable. The right plot compares the R Squared values of `Mkt.RF` + 3PC model and the Fama French 4 factor model (the `HML` is dropped), the R Squared value of the mixed PC model are significantly higher in the 4 banks' asset returns.

Approximate factor model via PCA In this approximation, the return data, either returns or excess returns are standardized, that is

each return series has 0 sample means and unit standard deviation. Thus, the factor model (8.7) is

$$(\mathbf{R}_t - \boldsymbol{\mu}_R) = \boldsymbol{B}^T \boldsymbol{F}_t + \boldsymbol{\varepsilon}_t\,. \tag{8.18}$$

For a fixed p, the principal component approximation of the covariance of the returns $\boldsymbol{R}_t$ is

$$\Sigma_R \approx \boldsymbol{O}_p \text{diag}\{\lambda_1, \ldots, \lambda_p\} \boldsymbol{O}_p^T + \Sigma_\varepsilon, \quad \boldsymbol{O}_p = (\boldsymbol{o}_1 \ldots, \boldsymbol{o}_p),$$

where λ_j and $\boldsymbol{o}_j$ are the jth eigenvalue and vector of Σ_R. The estimators of $\boldsymbol{O}_p$ and λ_j can obtained from the eigen decomposition of $\hat{\Sigma}_R$, the sample covariance of $\boldsymbol{R}_t$. This can easily be done by using R function `prcomp()`. For each t, the proxy for $\boldsymbol{F}_t$ is simply $\tilde{\boldsymbol{F}}_t = \hat{\boldsymbol{O}}_p^T \boldsymbol{R}_t$, which are available from output of `prcomp()`.

Because we work with the standardized $\boldsymbol{R}_t$, when fitting regression model (8.18) with proxy $\hat{\boldsymbol{F}}_t$, the estimated coefficient matrix $\hat{\boldsymbol{B}} = \hat{\boldsymbol{O}}_p^T$. That is, for each asset return R_{jt}, the estimated coefficients $\hat{\boldsymbol{\beta}}_j$ is the jth row vector of $\hat{\boldsymbol{O}}_p$. Furthermore, for each asset j, the R Squared is simply $\hat{\boldsymbol{\beta}}_j^T \hat{\Sigma}_{\hat{F}} \hat{\boldsymbol{\beta}}_j$, where $\hat{\Sigma}_{\hat{F}} = \text{diag}\{\hat{\lambda}_1, \ldots, \hat{\lambda}_p\}$.

The following R code is for $p = 3$ with excess returns of the 18 asset return in Eg. 8.1.

```
p = 3
Zt = apply(Yt, 2, function(u) (u-mean(u))/sd(u))
pca = prcomp(Zt)
B = t(pca$rotation[,1:p]) ## t(Op)
Ft.pc = pca$x[,1:p] ## proxy Ft
R.Sq.pc = diag(t(B)%*%diag(pca$sd[1:p]^2)%*%B)
```

Factor analysis

The goal of factor analysis is to extract unknown factor as well as their factor loadings $\boldsymbol{B}$'s from the data. The model is just like the factor model, however, in this approach, neither the factor values nor the loadings are directly observable. All that is available is the sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$. This is the same type of data available for PCA, 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 of $\boldsymbol{R}_t$ (or $\boldsymbol{Y}_t$) in (8.7) and its variance $\boldsymbol{\Sigma}_R = \mathbf{B}^T \boldsymbol{\Sigma}_F \boldsymbol{B} + \boldsymbol{\Sigma}_\varepsilon$ in (8.8). The only component of (8.8) that can be estimated directly from the data is $\boldsymbol{\Sigma}_R$. Perhaps this estimate can be used to find estimates of $\boldsymbol{B}$, $\boldsymbol{\Sigma}_F$ and $\boldsymbol{\Sigma}_\varepsilon$. However, without further constraints, it is not possible to identify all three matrices. The usual constraints are that the factors are uncorrelated and standardized, so that

$$\boldsymbol{\Sigma}_F = \mathrm{var}(\boldsymbol{F}_t) = \mathbf{I}_p .$$

With these constraints, the component of $\boldsymbol{\Sigma}_R$ simplifies to the statistical factor model

$$\boldsymbol{\Sigma}_R = \boldsymbol{B}^T \boldsymbol{B} + \boldsymbol{\Sigma}_\varepsilon . \tag{8.19}$$

Even with this simplification, $\boldsymbol{B}$ is only determined up to a rotation, that is, by multiplication with an orthogonal matrix. The model is indistinguishable from a model having $\tilde{\boldsymbol{B}} = \mathbf{P}\boldsymbol{B}$ where $\mathbf{P}$ is any

orthogonal matrix, $\mathbf{P}^T\mathbf{P} = \mathbf{I}_N$.

In order to determine $\boldsymbol{B}$, a further set of constraints is required. One possible set of constraints is that $\boldsymbol{B}\boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{B}^T$ is diagonal. Output from *R*'s function `facanal()` satisfies this constraint when the argument `rotation` is set to `none`. If $\boldsymbol{B}$ is rotated, then this constraint no longer holds. Under the assumption of normality, the estimates can be obtained by the MLE, which is the method used in R's `facanal()` function.

If the main purpose of the statistical factor model is to estimate $\boldsymbol{\Sigma}_R$ by (8.19), then the choice of constraint is irrelevant since all constraints lead to the same product $\boldsymbol{B}^T\boldsymbol{B}$. In particular, rotation of $\boldsymbol{B}$ does not change the estimate of $\boldsymbol{\Sigma}_R$.

Eg 8.5. This example continues the analysis of the 18 asset returns in Eg. 8.1. We will analyze the excess returns $\boldsymbol{Y}_t$ using the R function `factanal()`.

```
args(factanal)

## function (x, factors, data = NULL, covmat = NULL, n.obs = NA,
##     subset, na.action, start = NULL, scores = c("none", "regression",
##         "Bartlett"), rotation = "varimax", control = NULL, ...)
```

To fit a $p = 3$ factor model and with $\boldsymbol{B}\boldsymbol{\Sigma}_\varepsilon^{-1}\boldsymbol{B}^T$ being diagonal, we specify 3 and `none` for the arguments `factors` and `rotation`. Note that `factanal()` automatically standardizes the input data so that each variable has variance 1.

```r
fa.none = factanal(Yt,3, rotation = "none")
names(fa.none)

##  [1] "converged"    "loadings"     "uniquenesses" "correlation"
##  [5] "criteria"     "factors"      "dof"          "method"
##  [9] "STATISTIC"    "PVAL"         "n.obs"        "call"
```

The analysis output can be shown by applying `print()` of the output object.

```r
print(fa.none)

## Call:
## factanal(x = Yt, factors = 3, rotation = "none")
##
## Uniquenesses:
##  AAPL  MSFT  CSCO   IBM  INTC   CVX   XOM   EPD   PSX     C   BAC   JPM
## 0.491 0.396 0.456 0.544 0.616 0.183 0.152 0.532 0.353 0.146 0.089 0.120
##   WFC   MRK   ABT   BMY   LLY   PFE
## 0.242 0.686 0.496 0.756 0.706 0.684
##
## Loadings:
##      Factor1 Factor2 Factor3
## AAPL  0.511   0.493
## MSFT  0.536   0.555
## CSCO  0.600   0.428
## IBM   0.628   0.246
## INTC  0.503   0.359
## CVX   0.757           0.493
## XOM   0.738           0.546
## EPD   0.619           0.291
## PSX   0.710           0.374
## C     0.906  -0.106  -0.149
## BAC   0.916  -0.135  -0.232
## JPM   0.913          -0.198
## WFC   0.844  -0.126  -0.175
## MRK   0.430   0.359
## ABT   0.495   0.496  -0.113
## BMY   0.379   0.316
## LLY   0.329   0.431
## PFE   0.431   0.360
```



```
##
##                Factor1 Factor2 Factor3
## SS loadings      7.633   1.776   0.943
## Proportion Var   0.424   0.099   0.052
## Cumulative Var   0.424   0.523   0.575
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 1528.8 on 102 degrees of freedom.
## The p-value is 5.5e-253
```

The `Loadings` are the estimates of the $N \times p = 18 \times 3$ dimensional $\hat{B}^T$. Some of the elements are not shown because the default cutoff setting `cutoff = 0.1`, loading values with a modulus smaller than 0.1 are not printed. The print output gives the sums of squares of the 18 loadings for each factor. The `Proportion Var` row contains the `SS loadings` divided by $N = 18$, where $N = 18$ is the sum of the variances of the 18 variables, since each variable has been standardized to have variance equal to 1. The `uniqueness` are the diagonal elements of the estimate $\hat{\boldsymbol{\Sigma}}_\varepsilon$.

The first factor has all the same sign, it usually represents the general movement, the market component. The correlation between the first factor and `Mkt.RF` in the data is 0.828. The second factor has negative signs on the 4 banks. The third factor has negative signs on the 4 oil&gas companies and very small values (not printed) on the 4 banks. These two factors might be an industrial components. The loading of principle component factor model has similar properties The correlation between the first principal component and `Mkt.RF` is 0.900.

R's `factanal()` uses the MLE for a given number of factors p by treating F_t and ε_t are jointly normal, and thus $\boldsymbol{R}_t$ is normal with variance $\boldsymbol{B}^T\boldsymbol{B} + \boldsymbol{\Sigma}_\varepsilon$, where $\boldsymbol{\Sigma}_\varepsilon$ is diagonal. The likelihood ratio tests can be applied to check the adequacy of s fitted p-factor model. The test statistic is

$$\mathrm{LR}(p) = -\left(n - 1 - \frac{2N+5}{6} - \frac{2p}{3}\right)\left(\log\left|\hat{\boldsymbol{\Sigma}}_R\right| - \log\left|\hat{\boldsymbol{B}}^T\hat{\boldsymbol{B}} + \hat{\boldsymbol{\Sigma}}_\varepsilon\right|\right),$$

which, under the null hypothesis of p factors, is asymptotically distributed as a chi-squared distribution with degrees of freedom DF $= \frac{1}{2}\{(N-p)^2 - N - p\}$. This would require DF > 0, it follows that

$$p < \frac{1}{2}\left(2N + 1 - \sqrt{8N+1}\right). \tag{8.20}$$

This is also the upper bound of p in using `factanal()`.

The LR test statistic and its p-value are shown in the output, suggesting that p is greater than 3. Applying factor rotation sequentially with increasing p, the LR test is not significant when $p = 7$ for this data set.

Factor Rotation As mentioned earlier, $\boldsymbol{\Sigma}_R$, the factor model variance of $\boldsymbol{R}_t$ (8.19) is invariant to any rotation of $\boldsymbol{B}$. Rotation might increase the interpretability of the loadings, but there exist infinite possible factor rotations. There are many criteria proposed for factor rotation. The one available in R's `factanal()` is a varimax criterion of Kaiser (1958). In practice, factor rotation is used to aid the interpretations of common factors. It may be helpful in some applications, but not informative in others. In this example, we do not find the rotation with varimax criteria helpful in interpretability.

Approximate factor models via Factor Analysis With factor analysis, a candidate to proxy for the factor realizations is the generalized least squares regression estimator. Using the MLE of $\boldsymbol{B}$ and $\boldsymbol{\Sigma}_\varepsilon$, then for each t,

$$\hat{\boldsymbol{F}}_t = (\hat{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}\hat{\boldsymbol{B}}^T)^{-1}\hat{\boldsymbol{B}}\hat{\boldsymbol{\Sigma}}_\varepsilon^{-1}(\boldsymbol{R}_t - \bar{\boldsymbol{R}}). \tag{8.21}$$

This version of $\hat{\boldsymbol{F}}_t$ can be obtained from `factanal()` by setting `score = "Bartlett"`.

Eg 8.6. The data is the daily returns of 18 stocks from Eg. 8.1. We will work with the standardized excess returns $\boldsymbol{Y}_t$.

```
p = 3
Zt = apply(Yt, 2, function(u) (u-mean(u))/sd(u))
fa = factanal(Zt, p, scores = "Bartlett", rotation = "none")
B = t(fa$loading)
Ft.fa = fa$scores ## proxy Ft
R.Sq.fa = diag(t(B)%*%var(Ft.fa)%*%B)
```

The left panel of Figure 8.3 shows the R-squared values of the factor analysis and pca approximations comparing with those of the 3-factor Fama-French model in Eg. 8.1. Both approximations have higher R-squared values except the tech companies. We also fit the 3 models to the same data with 5 factors. The difference in R-squared values are more prominent. It appears that the factor analysis approximation perform the best.

Selecting the Number of Factors The remaining question in applying statistical factor models is how to select p the number of factors. Let $\hat{\sigma}_i^2(p)$ be the variance of residuals from the approximated

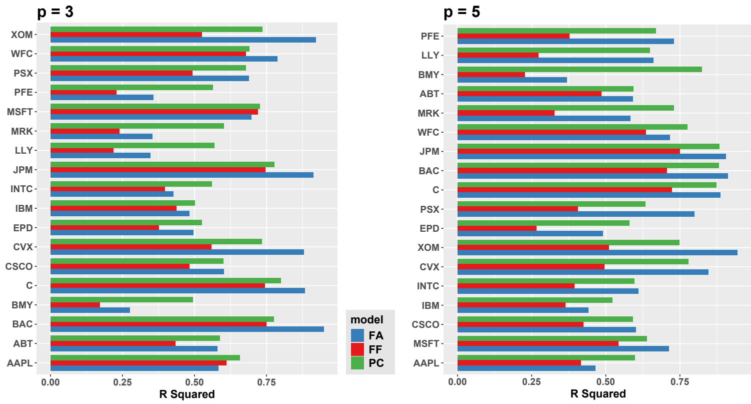


Figure 8.3: R-Squared values of fitting Fama-French model, factor analysis and principal component approximated models. The left panel uses 3 factors, the right panel 5 factors.

regression models with proxy of $\boldsymbol{F}_t$. Define

$$\hat{\sigma}^2(p) = \frac{1}{N}\sum_{i=1}^{N}\hat{\sigma}_i^2(p).$$

The criteria proposed by Bai and Ng (2002) adopts some information criteria and are defined as

$$C_1(p) = \hat{\sigma}^2(p) + p\hat{\sigma}^2(P)\left(\frac{N+n}{Nn}\right)\log\left(\frac{Nn}{N+n}\right),$$

$$C_2(p) = \hat{\sigma}^2(p) + p\hat{\sigma}^2(P)\left(\frac{N+n}{Nn}\right)\log N\ ,$$

where P is pre-specified positive integer denoting the maximum number of factors. We select the value of p that minimizes $C_1(\cdot)$ or $C_2(\cdot)$.

Appendix: Principal Component Analysis

Starting with a sample $\boldsymbol{Y}_i$, $i = 1, \ldots, n$, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iN})^T$ are N-dimensional random vector with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, principal components are linear combinations of random $Y_{i1}, \ldots, Y_{iN}$, which have special properties in terns of variances. For example, the first principal components is the normalized linear combination with maximum variance. The principal components turn out to be the eigen vectors of the covariance matrix.

One goal of **P**rincipal **C**omponent **A**nalysis is finding "structure" in $\boldsymbol{\Sigma}$. It can also be applied to either the sample covariance matrix or correlation matrix. We will use $\boldsymbol{\Sigma}$ to represent whichever matrix is chosen. We assume that the mean $\bar{\boldsymbol{Y}}$ has been subtracted from each $\boldsymbol{Y}_i$. The eigen-decomposition

$$\boldsymbol{\Sigma} = \boldsymbol{O}\mathrm{diag}(\lambda_1, \ldots, \lambda_N)\boldsymbol{O}^T,$$

where $\boldsymbol{O}$ is an orthogonal matrix, i.e., $\boldsymbol{O}^T\boldsymbol{O} = \boldsymbol{O}\boldsymbol{O}^T = \mathbf{I}_N$, whose columns $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_N$ are the eigenvectors corresponding eigenvalues $\lambda_1 > \ldots \lambda_N$ of $\boldsymbol{\Sigma}$. For convenience, we assume no ties among the eigenvalues.

The first principal component maximizes $\boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}$ over $\boldsymbol{a}$ such that $\boldsymbol{a}^T\boldsymbol{a} = 1$. The maximizer is $\boldsymbol{a} = \boldsymbol{o}_1$, the eigenvector corresponding to the largest eigenvalue, and is called the first principal axis. The projections, $\boldsymbol{o}_1^T\boldsymbol{Y}_i$, $i = 1, \ldots, n$ on this vector are first principal component. The set of eigen vectors, $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_N$ are the principal axes and the set of projections $\boldsymbol{o}_k^T\boldsymbol{Y}_i$, $i = 1, \ldots, n$ on to the kth eigenvector is the kth principal component.

The variance of the kth principal component is the kth eigenvalue,

$$\mathrm{var}(\boldsymbol{o}_k^T \boldsymbol{Y}_i) = \boldsymbol{o}_k^T \boldsymbol{\Sigma} \boldsymbol{o}_k = \lambda_k\,, \qquad i = 1, \dots, n\,.$$

Moreover, $\lambda_k/(\lambda_1 + \cdots \lambda_N)$ is the proportion of the variance due to this principal component and $(\lambda_1 + \cdots + \lambda_k)/(\lambda_1 + \cdots + \lambda_N)$ is the proportion of the variance due to the first k principal components. The principal components are mutually uncorrelated, that is, for $k, \ell = 1, \dots, N$ and $k \neq \ell$,

$$\mathrm{Cov}(\boldsymbol{o}_k^T \boldsymbol{Y}_i, \boldsymbol{o}_\ell^T \boldsymbol{Y}_i) = \boldsymbol{o}_k^T \boldsymbol{\Sigma} \boldsymbol{o}_\ell = 0\,, \qquad i = 1, \dots, n\,.$$

The matrix of principal components of original data is $\boldsymbol{S}$,

$$\boldsymbol{S} = \boldsymbol{Y}\boldsymbol{O}\,, \qquad \underset{n\times N}{\boldsymbol{Y}} = \left(\boldsymbol{Y}_1^T, \dots, \boldsymbol{Y}_n^T\right)^T, \quad \underset{N\times N}{\boldsymbol{O}} = \left(\boldsymbol{o}_1, \dots \boldsymbol{o}_N\right).$$

The matrix $\boldsymbol{S} = (\boldsymbol{S}_1^T, \dots, \boldsymbol{S}_n^T)^T$ and $\boldsymbol{S}_i$, $i = 1, \dots, n$ are N-dimensional random vectors $\boldsymbol{S}_i = (\boldsymbol{o}_1^T \boldsymbol{Y}_i, \dots, \boldsymbol{o}_N^T \boldsymbol{Y}_i)^T$. The eigenvectors are sometimes called the rotations as in output from R's `prcomp()` function.

In many applications, the first few principal components account for almost all of the variation, for most purposes, one can work solely with these principal components and discard the rest.

Variable Standardization If the components of $\boldsymbol{Y}_i$ have different units, such as dollars and rates, then it is necessary to standardize the variables so that those variables with larger magnitude would not completely dominate the PCA. This means we will work with

the correlation matrix instead of covariance matrix in such cases.

Eg 8.7. Daily yield of US Treasury bonds at 10 maturities, 1, 3, 6 moths, and 1, 2, 3, 5, 7, 10, and 20 years for the time period July 31, 2001 to October 22, 2018, $n = 4,310$.

```
head(yt,1)

##              m1   m3   m6   y1   y2   y3   y5   y7  y10  y20
## 2001-07-31 3.67 3.54 3.47 3.53 3.79 4.06 4.57 4.86 5.07 5.61

tail(yt,1)

##              m1   m3   m6   y1   y2   y3   y5   y7 y10  y20
## 2018-10-22 2.18 2.34 2.49 2.68 2.92 2.99 3.05 3.13 3.2 3.31
```

Figure 8.4 shows 3 component series, 3-month, 5-year and 20 year rates. The series appear to be nonstationary.

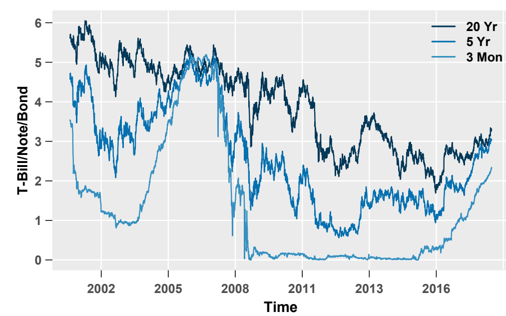


Figure 8.4: Daily time series plots of 3-month, 5-year and 20-year T-bill/note/bond rates from July 31,2001 to October 22, 2018.

Bond yields are higher for longer terms as shown in the left panel of Figure 8.5, the curve is the average yield over 4,310 daily yields. The right panel shows daily yield curves of 10 different dates. The daily yield curves can have a variety of shapes for different dates, some even have higher yields for shorter term T-Bills, as seen the yield curve of 2007-08-01.

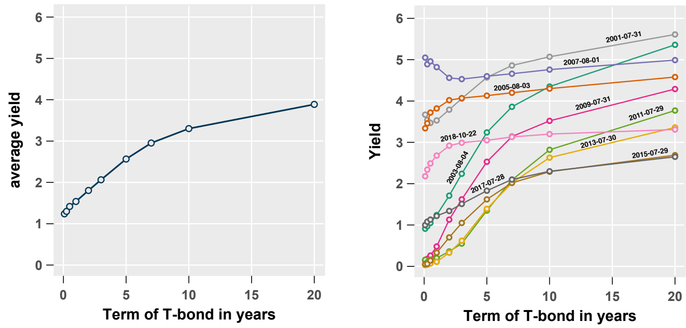


Figure 8.5: Average yield curve at the left panel and daily yield curves of 10 different dates at right panel.

To analyze daily changes in yield, we will study the differenced series, $\boldsymbol{y}_t - \boldsymbol{y}_{t-1}$, representing changes of the yield curves from day to day. Stationarity is also achieved by differencing as shown in Figure 8.6. These are based on the same 3 series 3-month, 5-year and 10 year rates in Figure 8.4.

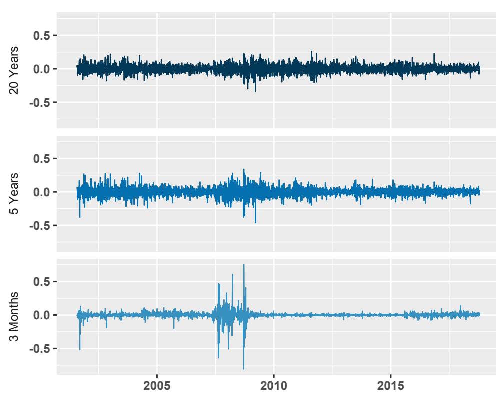


Figure 8.6: Differenced series plots of 3-month, 5-year and 20-year T-bill/note/bond rates from July 31,2001 to October 22, 2018.

The sample covariance matrix of differenced series is used because all 10 time series are in the same units. We will use R's function `prcomp()`.

```
pca = prcomp(dyt)
names(pca)

## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

The output value \$sdev gives $\sqrt{\hat{\lambda}_k}$ $k = 1, \ldots, N$; \$rotation is $\hat{\boldsymbol{O}}$ the principal axes or the eigen vactors; \$x is the matrix of principal components $\boldsymbol{S} = \boldsymbol{Y}\hat{\boldsymbol{O}}$. We can also print the summary.

```
summary(pca)

## Importance of components:
##                          PC1    PC2    PC3    PC4    PC5     PC6     PC7
## Standard deviation     0.139 0.0772 0.0426 0.0321 0.0207 0.01472 0.01188
## Proportion of Variance 0.664 0.2043 0.0622 0.0352 0.0147 0.00742 0.00484
## Cumulative Proportion  0.664 0.8679 0.9301 0.9654 0.9801 0.98752 0.99235
##                            PC8     PC9    PC10
## Standard deviation     0.01015 0.00807 0.00742
## Proportion of Variance 0.00353 0.00223 0.00189
## Cumulative Proportion  0.99588 0.99811 1.00000
```

The first row gives the values of $\sqrt{\hat{\lambda}_k}$, the second row the values of $\hat{\lambda}_k / \sum_{\ell=1}^{N} \hat{\lambda}_\ell$ and the third row the values of $\sum_{\ell=1}^{k} \hat{\lambda}_\ell / \sum_{\ell=1}^{N} \hat{\lambda}_\ell$ for $k = 1, \ldots, N = 10$. The first principle component account for 66.4% of the total variance. The first three principal components have 93.01% of the variance.

The left panel of Figure **??** plot the variances of the principal components. This type of plot is called a *scree plot*. The first free components have most of the variation.

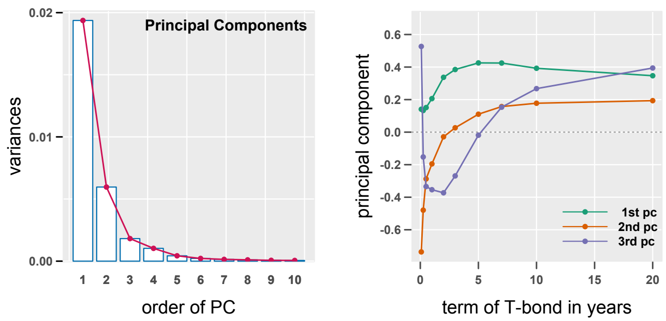


Figure 8.7: The left panel is the scree plot λ_i for the changes in T-bond yield. The right panel is the eigenvectors, $\hat{\boldsymbol{o}}_1$, $\hat{\boldsymbol{o}}_2$, $\hat{\boldsymbol{o}}_3$ corresponding to $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_3$.

The first component $\hat{\boldsymbol{o}}_1$ is a roughly equally weighted linear combination of the changes yield changes. This component might represent the general movement of the changes in interest rates. The second component has negative weights on the first 5 series and positive weights on the last 5 series, it represents the difference between the short and long term rates.

In constructing a bond portfolio, one would study the behavior of the yield changes overtime. In stead of analyzing the yields of all possible maturities, one can use the first two or three principal components instead of all 10 series. The cross correlations are 0 by construction.

We have used three principal components because they account for 93% of variance. Alternatively, an informal but useful procedure to determine the number of principal components needed in an application is to examine the scree plot, which is the plot of $\hat{\lambda}_i$. The number of components is chosen to be the number that appears prior to the *elbow*.

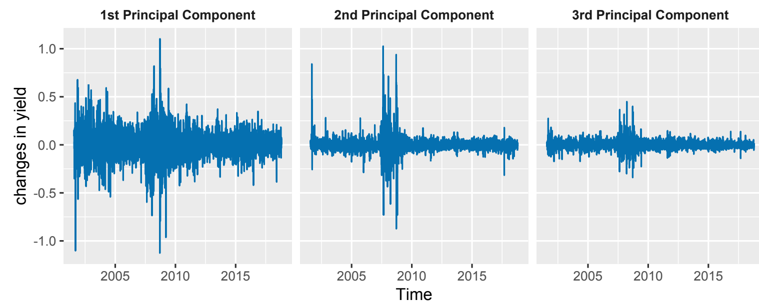


Figure 8.8: Time series plots of the first three principal components of the changes in yields of T-Bonds.

Selecting the first i principal components only provides an approximation to the total variance of the data. If a small i can provide a good approximation, then the simplification becomes valuable.