Stock Data Assignment

Soumya Mukherjee (M.Stat 1st Year)

22 May 2020

1 Description and preliminary analysis of the data

We are provided data on weekly gains in stock prices for 103 consecutive weeks for 5 companies namely Allied Chemical, Du Pont, Union Carbide, Exxon and Texaco. It is known that the first three are chemical companies and the last two are oil companies.

Let X_1, X_2, X_3, X_4 and X_5 denote the $p = 5$ random variables corresponding to the weekly gain in stock price corresponding to the companies Allied Chemical, Du Pont, Union Carbide, Exxon and Texaco, respectively which will be henceforth referred as 1^{st} Company, 2^{nd} Company and so on, for convenience. Let $\mathbf{x} = (X_1, X_2, X_3, X_4, X_5)^T$ be the 5 dimensional random variable of observable traits, in this case the vector of weekly gains in stock price for the 5 companies. Then we have $n = 103$ observations corresponding to these five random variables, or equivalently, $\mathbf{x}$. We store the data in the form of a $n \times p$ data matrix $\mathbf{M^{n\times p}}$. The first few and the last few rows of $\mathbf{M}$ is given by

$$\mathbf{M} = \begin{pmatrix} 0.0130338 & -0.0078431 & -0.0031889 & -0.0447693 & 0.0052151 \\ 0.0084862 & 0.0166886 & -0.0062100 & 0.0119560 & 0.0134890 \\ -0.0179153 & -0.0086393 & 0.0100360 & 0.0000000 & -0.0061428 \\ 0.0215589 & -0.0034858 & 0.0174353 & -0.0285917 & -0.0069534 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0.0033626 & 0.0029016 & -0.0030507 & -0.0012193 & -0.0097005 \\ 0.0170147 & 0.0095061 & 0.0181994 & -0.0161758 & -0.0075614 \\ 0.0103929 & -0.0026612 & 0.0044290 & -0.0024818 & -0.0164502 \\ -0.0127948 & -0.0143678 & -0.0187402 & -0.0049759 & -0.0163732 \end{pmatrix}$$

The sample mean vector is given by

$$\bar{x} = (0.0010627806, 0.0006554204, 0.0016260816, 0.0040491252, 0.0040386417)^T$$

The sample covariance matrix (with divisor $n - 1$) is given by

$$\mathbf{S} = \begin{pmatrix} 0.0004332695 & 0.0002756679 & 0.0001590265 & 0.0000641193 & 0.0000889662 \\ 0.0002756679 & 0.0004387172 & 0.0001799737 & 0.0001814512 & 0.0001232623 \\ 0.0001590265 & 0.0001799737 & 0.0002239722 & 0.0000734135 & 0.0000605461 \\ 0.0000641193 & 0.0001814512 & 0.0000734135 & 0.0007224964 & 0.0005082772 \\ 0.0000889662 & 0.0001232623 & 0.0000605461 & 0.0005082772 & 0.0007656742 \end{pmatrix}$$

The sample correlation matrix is given by

$$\mathbf{R} = \begin{pmatrix} 1.0000000 & 0.6322878 & 0.5104973 & 0.11460193 & 0.1544628 \\ 0.6322878 & 1.0000000 & 0.5741424 & 0.3222921 & 0.2126747 \\ 0.5104973 & 0.5741424 & 1.0000000 & 0.1824992 & 0.1462067 \\ 0.11460193 & 0.3222921 & 0.1824992 & 1.0000000 & 0.6833777 \\ 0.1544628 & 0.2126747 & 0.1462067 & 0.6833777 & 1.0000000 \end{pmatrix}$$

The pairwise scatterplots of observations corresponding to the 5 variables are as follows:

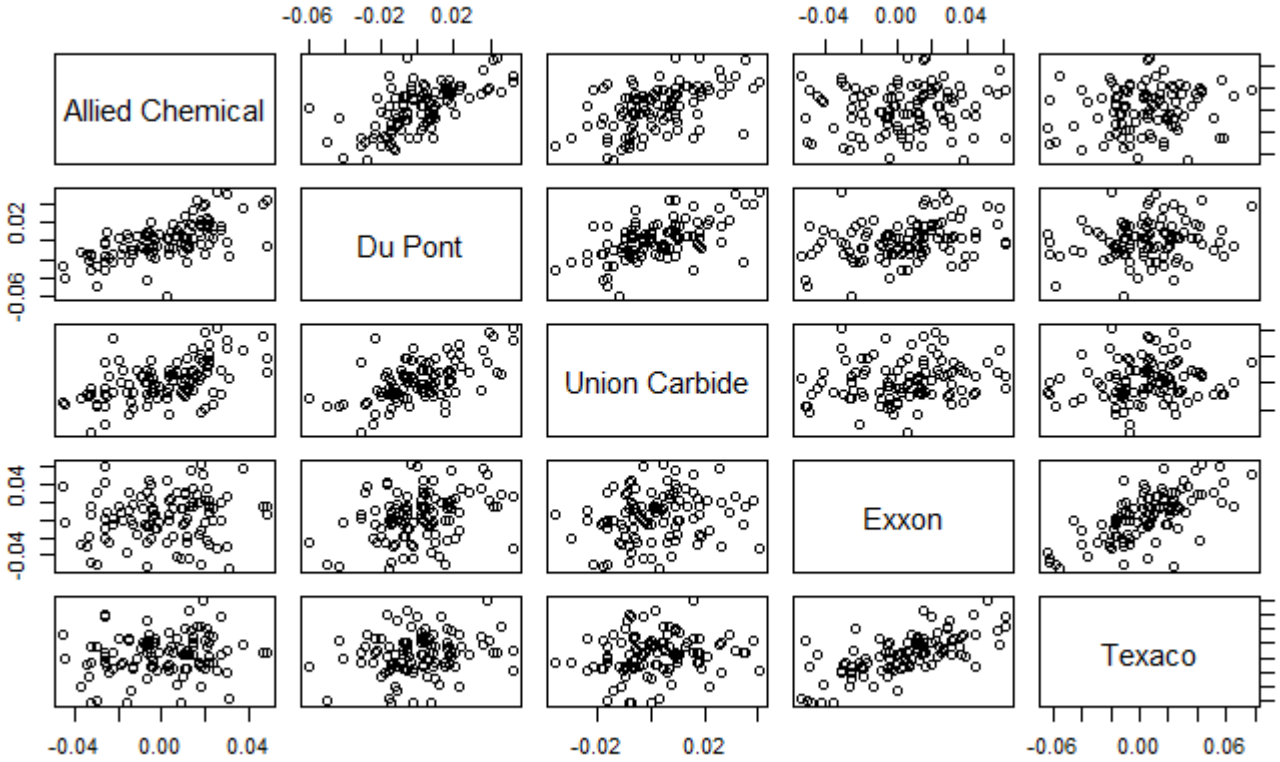


Figure 1: Pairwise scatterplots

To check whether the observations corresponding to the variables X_1, X_2, X_3, X_4 and X_5 can be considered as a random sample from some p-dimensional distribution, we perform the non-parametric Wald-Wolfowitz runs test of randomness [1] on each of the columns of the data matrix $\mathbf{M}$ and conclude that the null hypothesis of randomness of the sample is accepted at any reasonable level (5% or 10%) of significance for each of the variables.

Thus the observations in 103 successive weeks appear to be independently distributed, but weekly gain in stock prices across stocks are correlated, because as one expects, stocks tend to

move together in response to general economic conditions. Further, from the sample correlation matrix **R** as well as the pairwise scatterplots in Fig. 1, we observe that the first 3 companies Allied Chemical, Du Pont and Union Carbide have their weekly stock price gains highly correlated among themselves and the correlations are quite low with those corresponding to the last 2 companies Exxon and Texaco, while Exxon and Texaco too have their weekly stock price gains highly correlated among themselves. Further all sample correlations are positive. This suggests that all stock prices move together in general possibly due to a general "market factor". In addition prices for the chemical company stock prices move together and the same can be said for the oil company stock prices. This may be due to the effect of an underlying "industry factor" i.e. a factor that depends on whether the company is a chemical company or an oil company.

We further observe that all observed values are quite small and variability in the values in absolute terms is quite low which might lead to results which may not be quite useful or interpretable if we perform a factor analysis directly modelling the covariance matrix. However, variances of the weekly gains in stock prices are more or less same, so we can perform the factor analysis on the covariance matrix without expecting the factor loadings to be dominated by only a few variables with high variances.

2 Objective and Organization of the Report

Assuming an orthogonal factor model for these 5 variables, our objective is to estimate the unknown loading matrix, communalities and specific variances using Principal Components (PC), iterative PC and Maximum Likelihood (ML) method. We try to model separately the covariance as well as the correlation matrix. We then try to interpret the loadings of the variables in the factors. Then we perform varimax rotation on the loading matrix and see whether we can interpret the factors better. Finally we obtain the factor scores using both weighted least square and regression methods, and check whether they satisfy the assumptions of normality.

The report is organized as follows: **Section 1** gives the description of the stock price data along with conclusions based on a preliminary analysis of the data. **Section 2** outlines the objective of our study and the organization of this report. **Section 3** deals with checking the assumption of multivariate normality of the data, required for performing a large sample test for the number of common factors employed in our study. **Section 4** introduces the Orthogonal Factor Model that we fit to the covariance and correlation matrices. **Section 5** deals with determining the appropriate number of common factors to employed in case of the orthogonal factor models (OFM) for both the covariance and correlation matrices. **Section 6** is concerned with the estimation of the loading matrix and interpretation of the loadings corresponding to the separate orthogonal factor models (OFM) for both the covariance and correlation matrices. **Section 7** is concerned with the using the varimax rotation of the loading matrices and interpretation of the rotated loadings. **Section 8** gives the estimates of the factor scores using the weighted least squares and regression methods along with verification of the normality and independence assumptions made on the factor scores. **Section 9** gives a list of references used.

3 Checking the assumption of multivariate normality of the data

Although the assumption of multivariate normality of the data is not required for performing a factor analysis of the data and interpretation of the results, it is required for performing a large sample goodness of fit test for the number of common factors used in modelling the covariance or correlation matrices.

We want to determine whether the data follows a multivariate normal distribution. We utilize the fact that the marginal distribution of any component of a random vector which follows multivariate normal is a univariate normal for a preliminary test of multivariate normality. For this we use

- **Graphical Method :-** We plot the theoretical quantiles from the univariate normal distribution against the sample quantiles and obtain the QQ plot for each of the 5 variables X_1, X_2, X_3, X_4 and X_5. Using the asymptotic distribution of sample quantiles from the normal distribution, a 95% confidence belt is also constructed for each of the QQ plots. The QQ plots are shown in Fig. 2.

  Comments: From all the 5 QQ plots, we observe that there is appreciable agreement between sample and theoretical quantiles, although some observations are lying marginally outside the 95% confidence belt in the QQ plot for X_3. Hence we see no significant visual evidence to reject the hypothesis that the marginal distributions of all the 5 variables are univariate normal.

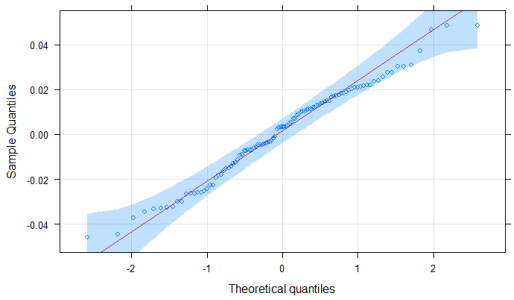
- **Formal testing procedure :-** To validate our observations from the QQ plots, we perform the Shapiro-Wilks Test for univariate normality for each of the 5 variables.

  The p-values of the Shapiro-Wilks tests are summarized in the following table:

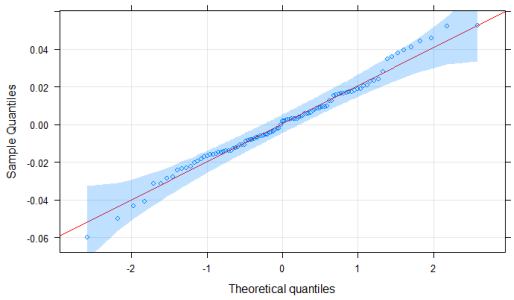
Variable	p-value
X_1	0.2222486
X_2	0.3512914
X_3	0.3719842
X_4	0.3994265
X_5	0.261365

  At any reasonable level of significance we can therefore conclude that the marginal distributions of the variables are univariate normal.

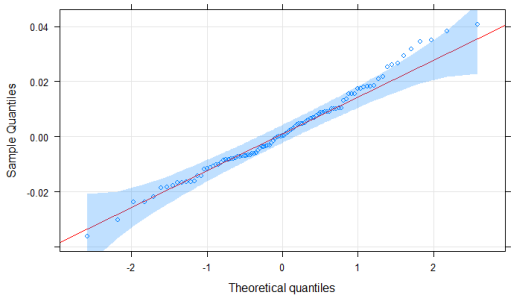
  However multivariate normality is not ensured even if the marginal distributions are univariate normal. To check whether the data follows a multivariate normal distribution, we use the multivariate extension of the Shapiro Wilks test of normality proposed by Villasenor-Alva and Gonzalez-Estrada [2], available in the **mvShapiroTest** package in R. The p-value of the test is obtained as 0.5194316. Hence the null hypothesis of multivariate normality of the data is accepted at any reasonable level of significance.



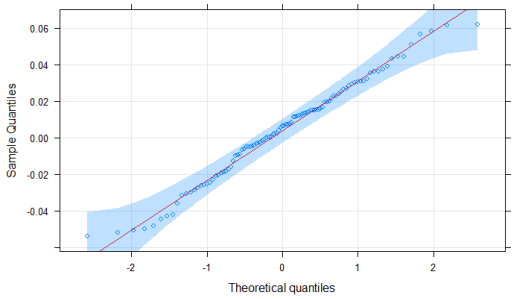
(a) QQ plot for X_1



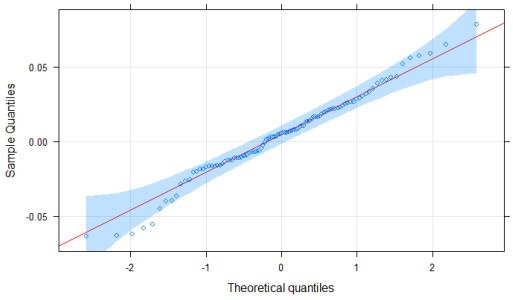
(b) QQ plot for X_2



(c) QQ plot for X_3



(d) QQ plot for X_4



(e) QQ plot for X_5

Figure 2: QQ plots of the 5 variables

4 The Orthogonal Factor Model

The development of the theory and notations used here largely follow Johnson and Wichern [3].

$\mathbf{X} = (X_1, \ldots, X_p)^T$ is a p-dimensional random vector of observable traits distributed with mean $\mu = (\mu_1, \mu_2, \ldots, \mu_p)^T$ and covariance matrix $\mathbf{\Sigma} = ((\sigma_{ij}))^{p\times p}$. Here $p = 5$.

The orthogonal factor model (OFM) assumes that $\mathbf{X}$ can be written as a linear combination of a set of common factors $F_1, F_2, \ldots, F_m$ and p additional unique factors $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$ (sometimes called specific factors).

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots + l_{1m}F_m + \epsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots + l_{2m}F_m + \epsilon_2 \\ &\vdots = \vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots + l_{pm}F_m + \epsilon_p \end{aligned}$$

where l_{ij} is the loading of the i -th variable on the j-th factor and the i-th specific factor ϵ_i is associated only with i -th variable X_i

The p deviations $X_1 - \mu_1, X_2 - \mu_2$ and $X_p - \mu_p$ are expressed in terms of $p + m$ unobservable random variables, $F_1, F_2, \ldots, F_m, \epsilon_1, \epsilon_2, \ldots, \epsilon_p$

In matrix notation,

$$\mathrm{X} - \boldsymbol{\mu} = \mathrm{L}_{p\times m}\mathrm{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1} \tag{1}$$

where $\mathbf{L}$ is the matrix of factor loadings and $\mathbf{F}$ is the vector of values of m unobservable variables (m common factors).

(1) is called an m -factor model.

The Orthogonal Factor Model (OFM) assumes

$E(\mathbf{F}) = \mathbf{0}$, $\quad \mathrm{Cov}(\mathbf{F}) = E\left(\mathbf{F}\mathbf{F}^T\right) = \mathbf{I}_m$ $\quad$ (Identity matrix of order m)

$E(\boldsymbol{\epsilon}) = \mathbf{0}$, $\quad \mathrm{Cov}(\boldsymbol{\epsilon}) = E\left(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right) = \Psi = \mathrm{diag}\left\{\psi_1, \psi_2, \ldots, \psi_p\right\}$ and $\mathrm{Cov}(\epsilon, \mathbf{F}) = \mathbf{0}_{p\times m}$

Hence the covariance structure of the m -factor model is given by

1. $\mathrm{Cov}(\mathrm{X}) = \mathbf{\Sigma} = \mathrm{L}\mathrm{L}^T + \Psi$ i.e.

$$\begin{aligned} \mathrm{Var}\left(X_i\right) &= l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i \\ \mathrm{Cov}\left(X_i, X_k\right) &= l_{i1}l_{k1} + \cdots + l_{im}l_{km} \end{aligned} \tag{2}$$

2. $\mathrm{Cov}(\mathrm{X}, \mathrm{F}) = \mathrm{L}$ i.e.

$$\mathrm{Cov}\left(X_i, F_k\right) = l_{ik} \tag{3}$$

The portion of $\sigma_{ii} = \mathrm{Var}\left(X_i\right)$ contributed by the m common factors is known as **communality**, denoted by h_i^2. The portion of σ_{ii} due to specific factor (not explained by common factors) is known as **uniqueness** or **specific variance**.

$$\underbrace{\sigma_{ii}}_{\text{Variance}} = \underbrace{l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2}_{\text{Communality} = h_i^2} + \underbrace{\psi_i}_{\text{Uniqueness}} \tag{4}$$

The i -th communality is the sum of squares of loadings of the i -th variable on the m common factors. The OFM assumes that $p(p+1)/2$ variances and covariances of X can be reproduced from the pm factor loadings and p specific variances.

When the variances corresponding to the variables $X_1, X_2, \ldots, X_p$ are vastly different, the loadings will typically be dominated by the variables having higher variability. In such cases, it is better to perform the factor analysis on the standardized variables

$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}},\ Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}},\ \ldots\ ,\ Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

In matrix notation,

$$\mathbf{Z} = \left(\mathbf{V}^{1/2}\right)^{-1} (\mathbf{X} - \boldsymbol{\mu})$$

where the diagonal standard deviation matrix $\mathbf{V}^{1/2}$ is defined as

$$\mathbf{V}^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

Clearly, $E(\mathbf{Z}) = \mathbf{0}$ and $\text{Cov}(\mathbf{Z}) = \left(\mathbf{V}^{1/2}\right)^{-1} \boldsymbol{\Sigma} \left(\mathbf{V}^{1/2}\right)^{-1} = \boldsymbol{\rho}$ where $\boldsymbol{\rho}$ is the population correlation matrix.

The covariance matrix $\boldsymbol{\rho}$ of $\mathbf{Z}$ has the representation

$$\boldsymbol{\rho} = \mathbf{V}^{-1/2} \boldsymbol{\Sigma} \mathbf{V}^{-1/2} = \left(\mathbf{V}^{-1/2}\mathbf{L}\right)\left(\mathbf{V}^{-1/2}\mathbf{L}\right)^T + \mathbf{V}^{-1/2} \Psi \mathbf{V}^{-1/2} \tag{5}$$

Thus, $\boldsymbol{\rho}$ has a factorization analogous to (2) with loading matrix $\mathbf{L_z} = (({}_z l_{ij}))^{p \times m} = \mathbf{V}^{-1/2}\mathbf{L}$ and specific variance matrix $\Psi_z = \text{diag}\left\{{}_z\psi_1,\ {}_z\psi_2, \ldots,\ {}_z\psi_p\right\} = \text{V}^{-1/2}\Psi\text{V}^{-1/2}$. This is exactly equivalent to directly imposing an OFM model as in (1) on the standardized variable $\mathbf{Z}$ instead of $\mathbf{X}$.

4.1 Orthogonal Rotation of the loading matrix

If m -factor model holds, then it also holds if the factors are rotated. If $\mathbf{G}$ is an $m \times m$ orthogonal matrix, then $\mathbf{X}$ can be written as

$$\mathbf{X} = \mu + \mathbf{L}\mathbf{G}\mathbf{G}^T\mathbf{F} + \boldsymbol{\epsilon}$$

since the random vector $\mathbf{G^T F}$ satisfies the conditions

$$E\left(\mathbf{G}^T\mathbf{F}\right) = \mathbf{0}, \quad \text{Var}\left(\mathbf{G}^T\mathbf{F}\right) = \mathbf{I}, \quad \text{Cov}\left(\mathbf{G}^T\mathbf{F}, \boldsymbol{\epsilon}\right) = \mathbf{0}$$

the m -factor model is valid with new factors $\mathbf{G^T F}$, loading matrix $\mathbf{LG}$ and $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{G}\mathbf{G}^T\mathbf{L}^T + \Psi$.

All communalities and specific variances remain unchanged due to this orthogonal rotation so this allows us to possibly rotate the loading matrix to be able able to interpret the factors in a better manner.

After deciding on the number of common factors m to be used in the model, The analysis of the factor model proceeds by imposing conditions that allow one to uniquely estimate $\mathbf{L}$ and Ψ. The loading matrix is then rotated (multiplied by an orthogonal matrix), where the rotation is determined by some "ease-of-interpretation" criterion. Here we will consider varimax rotation which will be discussed in due course.

5 Determining the appropriate number of common factors

The appropriate number of common factors to be used in the orthogonal factor model on the raw observations $\mathbf{X}$ or the standardized observations $\mathbf{Z}$ can be determined in the following ways:

1. Based on the Principal Component Method of estimation of the loading matrix using a residual matrix and examining proportion of variability explained by the factors.
2. Based on the Maximum Likelihood Method of estimation of the loading matrix using a large sample goodness of fit test.

The details of these two methods will be discussed under the sections dealing with the specific methods of estimation of the loading matrices.

6 Methods of estimation of the loading matrices

We will explore three methods of estimation of the loading matrices for the Orthogonal factor model corresponding to both the covariance and correlation matrices, namely:

1. The Principal Component(PC) Method
2. The Iterative Principal Component(PC) Method
3. The Maximum Likelihood(ML) Method

We will discuss these specific methods, their implementations and the interpretations of the factor loadings in the following 3 subsections.

6.1 The Principal Component(PC) Method of estimation of the loading matrices

First we discuss the modelling of the $\mathbf{\Sigma}$ matrix.

The spectral decomposition(SD) of $\mathbf{\Sigma}$ provides us with one factoring of the covariance matrix $\mathbf{\Sigma}$. Let $\mathbf{\Sigma}$ have eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{e}_i)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then

$$\begin{aligned}\mathbf{\Sigma} &= \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + \cdots + \lambda_p \mathbf{e}_p \mathbf{e}_p^T \\ &= \left[\sqrt{\lambda_1}\mathbf{e}_1 \sqrt{\lambda_2}\mathbf{e}_2 \cdots \sqrt{\lambda_p}\mathbf{e}_p\right] \begin{bmatrix} \sqrt{\lambda_1}\mathbf{e}_1^T \\ \sqrt{\lambda_2}\mathbf{e}_2^T \\ \vdots \\ \sqrt{\lambda_p}\mathbf{e}_p^T \end{bmatrix} \\ &= \mathbf{L}\mathbf{L}^{\mathbf{T}} + \mathbf{0}\end{aligned} \tag{6}$$

This fits the prescribed covariance structure for the factor analysis model having as many factors as variables ($m = p$) and specific variances $\psi_i = 0$ for all i. The loading matrix has j-th column given by $\sqrt{\lambda_j}\mathbf{e}_j$.

Although the factor analysis representation of $\mathbf{\Sigma}$ in (6) is exact, it is not particularly useful as it employs as many common factors as there are variables and does not allow for any variation in the specific factors $\boldsymbol{\epsilon}$ in (1). We prefer models that explain the covariance structure in terms of just a few common factors. We prefer models where $m > p$ such that the covariance structure can be explained using a small number of underlying factors.

One approach is that if, for some m, the last $p - m$ eigenvalues are small, we can ignore the last $p - m$ terms $\lambda_{m+1}\mathbf{e}_{m+1}\mathbf{e}_{m+1}^T + \ldots + \lambda_p \mathbf{e}_p \mathbf{e}_p^T$ in the spectral decomposition(SD) of Σ

$$\mathbf{\Sigma} \approx \mathbf{L}_{p\times m}\mathbf{L}_{m\times p}^T = \left[\sqrt{\lambda}_1\mathbf{e}_1 \sqrt{\lambda}_2\mathbf{e}_2 \ldots \sqrt{\lambda}_m\mathbf{e}_m\right] \begin{bmatrix} \sqrt{\lambda}_1\mathbf{e}_1^T \\ \sqrt{\lambda}_2\mathbf{e}_2^T \\ \vdots \\ \sqrt{\lambda}_m\mathbf{e}_m^T \end{bmatrix} \tag{7}$$

Then the underlying assumption is that the specific factor $\boldsymbol{\epsilon}$ is of minor importance and can be ignored in the factoring of $\mathbf{\Sigma}$.

The communality for the $i-$ th observed variable is the amount of its variance that can be attributed to the variation in the m factors, $h_i^2 = \sum_{j=1}^m l_{ij}^2$

If specific factors are included in the model, then their variances can be taken to be the diagonal elements of the difference matrix $\Sigma - \mathrm{L}_{p\times m}\mathrm{L}_{m\times p}^T$ and

$$\Psi = \operatorname{diag}\left\{\psi_1, \ldots, \psi_p\right\}, \quad \psi_i = \sigma_{ii} - \sum_{j=1}^m l_{ij}^2, \quad i = 1, \ldots, p; \quad \Psi = \operatorname{diag}\left(\mathbf{\Sigma} - \mathbf{L}_{p\times m}\mathbf{L}_{m\times p}^T\right) \tag{8}$$

Note that using $m < p$ factors will produce an approximation

$$\mathbf{L}_{p\times m}\mathbf{L}_{m\times p}^T + \Psi$$

for $\mathbf{\Sigma}$ that exactly reproduces the variances of the p measured variables but only approximates the covariances/correlations.

If variables are measured in very different scales, we work with the standardized variables as is done when extracting principal components. This is equivalent to modeling the correlation matrix ρ rather than the covariance matrix $\mathbf{\Sigma}$.

6.1.1 The PC method

In practice, to apply this approach to the data set at hand, we work with the sample covariance matrix $\mathbf{S} = ((s_{ij}))^{p\times p}$ in place of $\mathbf{\Sigma}$.

The principal component factor analysis of the sample covariance matrix $\mathbf{S}$ is specified in terms of its eigenvalue-eigenvector pairs $\left(\hat{\lambda}_1, \hat{\mathbf{e}}_1\right), \left(\hat{\lambda}_2, \hat{\mathbf{e}}_2\right), \ldots \left(\hat{\lambda}_p, \hat{\mathbf{e}}_\mathrm{p}\right)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. Let $m < p$ be the number of common factors. Then the matrix of estimated factor loadings $\left\{\tilde{\ell}_{ij}\right\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1; \sqrt{\hat{\lambda}_2}\hat{\mathbf{e}}_2 : \cdots ; \sqrt{\hat{\lambda}_m}\hat{\mathbf{e}}_\mathrm{m}\right] \tag{9}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S} - \widetilde{\mathbf{L}}\tilde{\mathbf{L}}^T$, so

$$\tilde{\mathbf{\Psi}} = \begin{bmatrix} \widetilde{\psi}_\mathbf{1} & 0 & \cdots & 0 \\ 0 & \widetilde{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{\psi}_p \end{bmatrix} \quad \text{with} \quad \tilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \tilde{\ell}_{ij}^2 \tag{10}$$

Communalities are estimated as

$$\tilde{h}_i^2 = \tilde{\ell}_{i1}^2 + \widetilde{\ell}_{i2}^2 + \cdots + \tilde{\ell}_{im}^2 \tag{11}$$

The principal component factor analysis of the sample correlation matrix is obtained by starting with $\mathbf{R}$ in place of $\mathbf{S}$. All procedures are analogous except for the replacement of s_{ii} by 1 in (10).

6.1.2 Choice of the number of common factors m

One method of choosing m can be based on the estimated eigenvalues in much the same manner as with principal components. Consider the residual matrix

$$\mathbf{S} - \left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T + \widetilde{\Psi}\right) \tag{12}$$

resulting from the approximation of $\mathbf{S}$ by the principal component solution. The diagonal elements are zero, and if the other elements are also small, we may subjectively take the m factor model to be appropriate. Analytically, we have

$$\text{Sum of squared entries of } \left(\mathbf{S} - \left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T + \widetilde{\Psi}\right)\right) \leq \hat{\lambda}_{m+1}^2 + \cdots + \hat{\lambda}_p^2$$

Consequently, a small value for the sum of the squares of the neglected eigenvalues implies a small value for the sum of the squared errors of approximation.

Ideally, the contributions of the first few factors to the sample variances of the variables should be large. The contribution to the sample variance s_{ii} from the first common factor is $\widetilde{\ell}_{i1}^2$. The contribution to the total sample variance, $s_{11}+ s_{22} + \cdots + s_{pp} = \text{tr}(\mathbf{S})$ from the first common factor is then

$$\widetilde{\ell}_{11}^2 + \widetilde{\ell}_{21}^2 + \cdots + \widetilde{\ell}_{p1}^2 = \left(\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1\right)^T \left(\sqrt{\hat{\lambda}_1}\hat{\mathbf{e}}_1\right) = \hat{\lambda}_1$$

since the eigenvector $\hat{\mathbf{e}}_1$ has unit length.

In general,

$$\left(\begin{array}{c}\text{Proportion of total}\\ \text{sample variance}\\ \text{due to j-th factor}\end{array}\right) = \begin{cases} \frac{\hat{\lambda}_j}{s_{11}+s_{22}+\cdots+s_{pp}} & \text{for a factor analysis of } \mathbf{S} \\ \frac{\hat{\lambda}_j}{p} & \text{for a factor analysis of } \mathbf{R} \end{cases}$$

The above criterion is frequently used as a heuristic device for determining the appropriate number of common factors. The number of common factors retained in the model is increased until a "suitable proportion" of the total sample variance has been explained.

Another convention, frequently encountered in packaged computer programs, is to set m equal to the number of eigenvalues of $\mathbf{R}$ greater than one if the sample correlation matrix is factored, or equal to the number of positive eigenvalues of $\mathbf{S}$ if the sample covariance matrix is factored. These rules of thumb should not be applied indiscriminately. For example, $m = p$ if the rule for $\mathbf{S}$ is obeyed, since all the eigenvalues are expected to be positive for large sample sizes. The best approach is to retain few rather than many factors, assuming that they provide a satisfactory interpretation of the data and yield a satisfactory fit to $\mathbf{S}$ or $\mathbf{R}$.

6.1.3 Application of the PC method to stock price data

The sample covariance matrix $\mathbf{S}$ and the sample correlation matrix $\mathbf{R}$ are as obtained in **Section 1** of this report. The eigenvalues and eigenvectors of $\mathbf{S}$ and $\mathbf{R}$ are as follows:

Eigenvalues of S:

$$\begin{bmatrix}\hat{\lambda}_1 & \hat{\lambda}_2 & \hat{\lambda}_3 & \hat{\lambda}_4 & \hat{\lambda}_5\end{bmatrix} = \begin{bmatrix}0.0013676780 & 0.0007011596 & 0.0002538024 & 0.0001426026 & 0.0001188868\end{bmatrix}$$

Matrix of Eigenvectors of S:

$$P = \begin{bmatrix}\hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 & \hat{\mathbf{e}}_4 & \hat{\mathbf{e}}_5\end{bmatrix} = \begin{bmatrix} -0.2228228 & 0.6252260 & -0.32611218 & 0.6627590 & -0.11765952 \\ -0.3072900 & 0.5703900 & 0.24959014 & -0.4140935 & 0.58860803 \\ -0.1548103 & 0.3445049 & 0.03763929 & -0.4970499 & -0.78030428 \\ -0.6389680 & -0.2479475 & 0.64249741 & 0.3088689 & -0.14845546 \\ -0.6509044 & -0.3218478 & -0.64586064 & -0.2163758 & 0.09371777 \end{bmatrix}$$

Eigenvalues of R:

$$\begin{bmatrix} {}_z\hat{\lambda}_1 & {}_z\hat{\lambda}_2 & {}_z\hat{\lambda}_3 & {}_z\hat{\lambda}_4 & {}_z\hat{\lambda}_5 \end{bmatrix} = \begin{bmatrix} 2.4372731 & 1.4070127 & 0.5005127 & 0.4000316 & 0.2551699 \end{bmatrix}$$

Matrix of Eigenvectors of R:

$$P_z = \begin{bmatrix} {}_z\hat{\mathbf{e}}_1 & {}_z\hat{\mathbf{e}}_2 & {}_z\hat{\mathbf{e}}_3 & {}_z\hat{\mathbf{e}}_4 & {}_z\hat{\mathbf{e}}_5 \end{bmatrix} = \begin{bmatrix} -0.4690832 & 0.3680070 & -0.60431522 & 0.3630228 & 0.38412160 \\ -0.5324055 & 0.2364624 & -0.13610618 & -0.6292079 & -0.49618794 \\ -0.4651633 & 0.3151795 & 0.77182810 & 0.2889658 & 0.07116948 \\ -0.3873459 & -0.5850373 & 0.09336192 & -0.3812515 & 0.59466408 \\ -0.3606821 & -0.6058463 & -0.10882629 & 0.4934145 & -0.49755167 \end{bmatrix}$$

The proportions of total variance explained by the principal components in case of both **S** and **R** are as follows:

Principal Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1st PC	$\hat{\lambda}_1$=0.0013676780	52.926066	52.92607
2nd PC	$\hat{\lambda}_2$=0.0007011596	27.133298	80.05936
3rd PC	$\hat{\lambda}_3$=0.0002538024	9.821584	89.88095
4th PC	$\hat{\lambda}_4$=0.0001426026	5.518400	95.39935
5th PC	$\hat{\lambda}_5$=0.0001188868	4.600652	100.00000

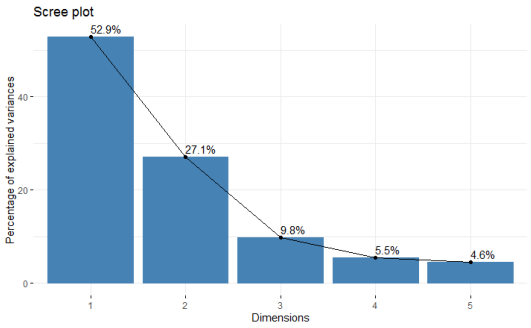
Table 1: Percentage of total variation explained by principal components of **S**

Principal Component	Eigenvalue	Percentage of variance	Cumulative percentage of variance
1st PC	${}_z\hat{\lambda}_1$=2.4372731	48.745462	48.74546
2nd PC	${}_z\hat{\lambda}_2$=1.4070127	28.140253	76.88572
3rd PC	${}_z\hat{\lambda}_3$=0.5005127	10.010255	86.89597
4th PC	${}_z\hat{\lambda}_4$=0.4000316	8.000632	94.89660
5th PC	${}_z\hat{\lambda}_5$=0.2551699	5.103398	100.00000

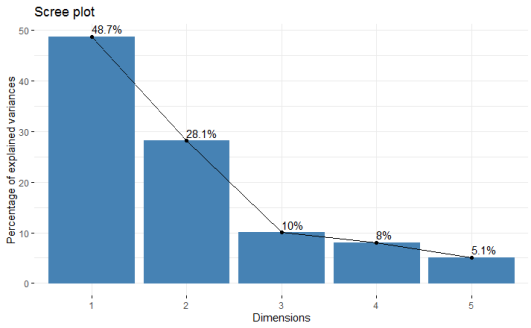
Table 2: Percentage of total variation explained by principal components of **R**

The Scree plots displaying the percentages of variation explained by the principal components graphically are as follows:

Comments: From the percentage of total variance explained by the principal components as shown in Tables 1 and 2 we find that approximately 80% of the total variance is explained by the 1st two principal components in case of **S**, while in case of **R**, it is almost 77%. Thus a choice of employing $m = 2$ common factors seems to be appropriate in both cases. This observation is supported by the scree plots obtained in Fig. 3 where in both the plots a sharp elbow is visible as one moves from the 2^{nd} to the 3^{rd} principal component.



(a) Scree Plot for **S**



(b) Scree Plot for **R**

Figure 3: Scree Plots

So we perform the factor analysis of both **S** and **R** employing $m = 2$ common factors.

Let $\widetilde{L} = \begin{bmatrix} \widetilde{L}_1 & \widetilde{L}_2 \end{bmatrix}$ and ${}_z\widetilde{L} = \begin{bmatrix} {}_z\widetilde{L}_1 & {}_z\widetilde{L}_2 \end{bmatrix}$ be the estimates of the loading matrices corrseponding to the OFM imposed on **S** and **R** respectively.

The estimated factor loadings, communalities and specific variances using the PC method are obtained as shown in Tables 3 and 4:

Variable	Estimated Factor Loadings		Communalities	Specific variances
	$\tilde{L_1}$	$\tilde{L_2}$	$\tilde{h}_i^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$
X_1(Allied Chemical)	-0.008240463	0.016555621	0.0003419938	0.00009127563
X_2(Du Pont)	-0.011364238	0.015103596	0.0003572645	0.00008145269
X_3(Union Carbide)	-0.005725215	0.009122290	0.0001159943	0.0001079779
X_4(Exxon)	-0.023630398	-0.006565506	0.0006015016	0.0001209948
X_5(Texaco)	-0.024071832	-0.006565506	0.0006520834	0.0001135907

Table 3: Estimates using PC method for **S**

Comments/Interpretation of factors:

From the factor loadings corresponding to the covariance matrix **S** in Table 3, we see that all the variable loadings are close to zero. So we conclude that interpretations based on these factor loadings are not useful. We do observe though that all the variables have the same sign in the first factor loading while the variables corresponding to the chemical and oil companies have different signs in the second factor loading. This might indicate some structure in the covariance matrix which may be further revealed by the study of the correlation matrix **R**.

Now we proceed to interpret the factor loadings obtained by modelling **R**. It seems fairly

Variable	Estimated Factor Loadings		Communalities	Specific variances
	${}_z\tilde{L}_1$	${}_z\tilde{L}_2$	${}_z\tilde{h}_i^2$	${}_z\tilde{\psi}_i = 1 - {}_z\tilde{h}_i^2$
X_1(Allied Chemical)	-0.7323218	0.4365209	0.7268458	0.2731542
X_2(Du Pont)	-0.831179	0.2804859	0.7695311	0.2304689
X_3(Union Carbide)	-0.7262022	0.3738582	0.6671396	0.3328604
X_4(Exxon)	-0.6047155	-0.6939569	0.8472571	0.1527429
X_5(Texaco)	-0.5630885	-0.7186401	0.8335122	0.1664878

Table 4: Estimates using PC method for $\mathbf{R}$

clear from the factor loadings in Table 4 that the first factor, ${}_z\tilde{L}_1$, represents general economic conditions and might be called a market factor. All of the stocks load highly on this factor, and the loadings are about equal. The second factor contrasts the chemical stocks with the oil stocks. (The chemical companies have relatively large negative loadings, and the oil companies have large positive loadings, on the factor.) Thus, ${}_z\tilde{L}_2$ seems to differentiate stocks in different industries and might be called an industry factor. The first factor appears to be a market-wide effect on weekly stock price gains whereas the second factor reflects industry specific effects on chemical and oil stock price returns. To summarize, weekly stock price gains appear to be determined by general market conditions and activities that are unique to the different industries, as well as a residual or company specific factor.

An idea of how well the estimated OFM model fits the covariance or correlation matrix is obtained by calculating the residual matrices $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathbf{T}} - \tilde{\mathbf{\Psi}}$ and $\mathbf{R} - {}_{\mathbf{z}}\tilde{\mathbf{L}}_{\mathbf{z}}\tilde{\mathbf{L}}^{\mathbf{T}} - {}_{\boldsymbol{z}}\tilde{\mathbf{\Psi}}$ (symbols represent obvious meanings) and observing whether the off-diagonal elements of the residual matrices are sufficiently close to zero.

$$\mathbf{S}-\left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T+\widetilde{\Psi}\right)=\begin{pmatrix} 0.00000000000 & -0.00006802809 & -0.00003917707 & -0.00002191010 & 0.00003169578 \\ -0.00006802809 & 0.00000000000 & -0.00002286839 & 0.00001207249 & -0.00002157775 \\ -0.00003917707 & -0.00002286839 & 0.00000000000 & -0.00000198318 & 0.00000047298 \\ -0.00002191010 & 0.00001207249 & -0.00000198318 & 0.00000000000 & -0.00011650330 \\ 0.00003169578 & -0.00002157775 & 0.00000047298 & -0.00011650330 & 0.00000000000 \end{pmatrix}$$

$$\mathbf{R}-\left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^T+{}_z\widetilde{\Psi}\right)=\begin{pmatrix} 0.00000000 & -0.09884073 & -0.184513342 & -0.025317756 & 0.055802201 \\ -0.09884073 & 0.00000000 & -0.134323655 & 0.014310342 & -0.053784258 \\ -0.18451334 & -0.13432365 & 0.000000000 & 0.002794963 & 0.005960127 \\ -0.02531776 & 0.01431034 & 0.002794963 & 0.000000000 & -0.155835955 \\ 0.05580220 & -0.05378426 & 0.005960127 & -0.155835955 & 0.000000000 \end{pmatrix}$$

The off-diagonal elements of $\mathbf{S}-\left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T+\widetilde{\Psi}\right)$ are all very close to zero as expected. However the (1,3), (2,3) and (4,5) positions (as well the symmetric counterparts) of the matrix given by $\mathbf{R}-\left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^T+{}_z\widetilde{\Psi}\right)$ are quite different from 0. Since this might indicate that the number of

common factors used is not appropriate, we also increased the number of common factors to 3, but there was no significant improvement.

6.2 The Iterative Principal Component(PC) Method of estimation of the loading matrices

A modification of the principal component approach is the iterative principal component method which allows for greater focus on explaining the covariances or correlations. The reasoning here is described in terms of a factor analysis of $\mathbf{R}$, although the procedure is also appropriate for $\mathbf{S}$.

If the factor model $\boldsymbol{\rho} = ((\rho_{ij}))^{p\times p} = \mathbf{L}\mathbf{L}^T + \Psi$ is correctly specified, the m common factors should account for the off-diagonal elements of $\boldsymbol{\rho}$, as well as the communality portions of the diagonal elements

$$\rho_{ii} = 1 = h_i^2 + \psi_i$$

If the specific factor contribution ψ_i is removed from the diagonal or, equivalently, the 1 replaced by h_i^2, the resulting matrix is $\boldsymbol{\rho} - \boldsymbol{\Psi} = \mathbf{L}\mathbf{L}^T$

Suppose, now, that initial estimates ψ_i^* of the specific variances are available. Then replacing the i-th diagonal element of $\mathbf{R}$ by $h_i^{*2} = 1 - \psi_i^*$, we obtain a "reduced" sample correlation matrix

$$\mathbf{R}_r = \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & h_p^{*2} \end{bmatrix} \tag{13}$$

Now, apart from sampling variation, all of the elements of the reduced sample correlation matrix $\mathbf{R}_r$ should be accounted for by the m common factors. In particular, $\mathbf{R}_r$ is factored as

$$\mathbf{R}_r = \mathbf{L}_r^* \mathbf{L}_r^{*T} \tag{14}$$

where $\mathbf{L}_r^* = \left\{\ell_{ij}^*\right\}$ are the estimated loadings.

The iterative principal component method of factor analysis employs the estimates

$$\begin{aligned} \mathbf{L}_r^* &= \left[\sqrt{\hat{\lambda}_1^*}\hat{\mathbf{e}}_1^* \quad \sqrt{\hat{\lambda}_2^*}\hat{\mathbf{e}}_2^* \quad \cdots \quad \sqrt{\hat{\lambda}_m^*}\hat{\mathbf{e}}_m^*\right] \\ \psi_i^* &= 1 - \textstyle\sum_{j=1}^m \ell_{ij}^{*2} \end{aligned} \tag{15}$$

where $\left(\hat{\lambda}_i^*, \hat{\mathbf{e}}_i^*\right), i = 1, 2, \ldots, m$ are the (largest) eigenvalue-eigenvector pairs determined from $\mathbf{R}_r$. In turn, the communalities would then be re-estimated by

$$\tilde{h}_i^{*2} = \sum_{j=1}^m \ell_{ij}^{*2} \tag{16}$$

The principal factor solution can be obtained iteratively, with the communality estimates obtained above in Eq.(16) becoming the initial estimates for the next stage.

In the spirit of the principal component solution, consideration of the estimated eigenvalues $\hat{\lambda}_1^*, \hat{\lambda}_2^*, \ldots, \hat{\lambda}_p^*$ helps determine the number of common factors to retain. An added complication

is that now some of the eigenvalues may be negative, due to the use of initial communality estimates. Ideally, we should take the number of common factors equal to the rank of the reduced population matrix. Unfortunately, this rank is not always well determined from $\mathbf{R}_r$, and some judgment is necessary.

Although there are many choices for initial estimates of specific variances, the most popular choice, when one is working with a correlation matrix, is $\psi_i^* = 1/r^{ii}$ where r^{ii} is the i th diagonal element of $\mathbf{R}^{-1}$. The initial communality estimates then become

$$h_i^{*2} = 1 - \psi_i^* = 1 - \frac{1}{r^{ii}}$$

which is equal to the square of the multiple correlation coefficient between X_i and the other $p-1$ variables. The relation to the multiple correlation coefficient means that h_i^{*2} can be calculated even when $\mathbf{R}$ is not of full rank. For factoring $\mathbf{S}$, the initial specific variance estimates use s^{ii}, the diagonal elements of $\mathbf{S}^{-1}$.

The Iterative principal component algorithm is thus outlined as follows:

The Iterative Principal Component(PC) Algorithm

---

1. Let ψ_i^* be the initial estimates of the specific variances.
2. Obtain initial estimates of the communalities $\tilde{h}_1^{*2}, \tilde{h}_2^{*2}, \ldots, \tilde{h}_p^{*2}$ using Eq.(14) , Eq.(15) and Eq.(16).
3. Estimate loading matrix

$$\mathbf{R} - \mathbf{\Psi}^* \approx \begin{bmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{12} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & h_p^{*2} \end{bmatrix}$$

   where $h_i^{*2} = 1 - \psi_i^*$.
4. Use spectral decomposition of $\mathbf{R} - \mathbf{\Psi}^*$ to find a good approximation of R_r.
5. The estimated loading matrix is obtained from the eigenvalue-eigenvector pairs $\left(\hat{\lambda}_i^*, \hat{\mathbf{e}}_i^*\right), i = 1, 2, \ldots, m$ of the matrix R_r

$$\mathbf{L}_r^* = \left[\sqrt{\hat{\lambda}_1^*}\hat{\mathbf{e}}_1^* \quad \sqrt{\hat{\lambda}_2^*}\hat{\mathbf{e}}_2^* \quad \cdots \quad \sqrt{\hat{\lambda}_m^*}\hat{\mathbf{e}}_m^*\right] = ((\ell_{ij}^*))^{p\times m}$$

6. Update specific variances as $\psi_i^* = 1 - \sum_{j=1}^m \ell_{ij}^{*2}$.
7. Repeat until some convergence criterion is achieved.

6.2.1 Application of the Iterative PC method to stock price data

The sample covariance matrix **S** and the sample correlation matrix **R** are as obtained in **Section 1** of this report. As in the application of the Principal Component(PC) method, we perform the factor analysis of both **S** and **R** employing $m = 2$ common factors. We used a maximum of 50 iterations.

Let $\widetilde{L} = \begin{bmatrix} \widetilde{L}_1 & \widetilde{L}_2 \end{bmatrix}$ and ${}_z\widetilde{L} = \begin{bmatrix} {}_z\widetilde{L}_1 & {}_z\widetilde{L}_2 \end{bmatrix}$ be the estimates of the loading matrices corresponding to the OFM imposed on **S** and **R** respectively.

The estimated factor loadings, communalities and specific variances using the Iterative PC method are as follows:

Variable	Estimated Factor Loadings		Communalities	Specific variances
	$\tilde{L_1}$	$\tilde{L_2}$	$\tilde{h}_i^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$
X_1(Allied Chemical)	0.008812028	0.012278722	0.0002284188	0.0002048506
X_2(Du Pont)	0.012049919	0.011312633	0.0002731762	0.0001655410
X_3(Union Carbide)	0.006536749	0.007934450	0.0001056846	0.0001182876
X_4(Exxon)	0.019842768	-0.007255823	0.0004463824	0.0002761140
X_5(Texaco)	0.019231993	-0.007924641	0.0004326695	0.0003330047

Table 5: Estimates using Iterative PC method for **S**

Variable	Estimated Factor Loadings		Communalities	Specific variances
	${}_z\tilde{L}_1$	${}_z\tilde{L}_2$	${}_z\tilde{h}_i^2$	${}_z\tilde{\psi}_i = 1 - {}_z\tilde{h}_i^2$
X_1(Allied Chemical)	0.6254584	-0.4293152	0.5755098	0.424490154
X_2(Du Pont)	0.7766148	-0.3417133	0.7198985	0.280101509
X_3(Union Carbide)	0.5909665	-0.3320045	0.4594683	0.540531664
X_4(Exxon)	0.7035483	0.7087661	0.9973296	0.002670401
X_5(Texaco)	0.5087849	0.4549090	0.4658042	0.534195813

Table 6: Estimates using Iterative PC method for **R**

Comments/Interpretation of factors:

From the factor loadings corresponding to the covariance matrix **S** in Table 5, we see that all the variable loadings are close to zero. So we conclude that interpretations based on these

factor loadings are not useful. We do observe though that all the variables have the same sign in the first factor loading while the variables corresponding to the chemical and oil companies have different signs in the second factor loading. This might indicate some structure in the covariance matrix which may be further revealed by the study of the correlation matrix $\mathbf{R}$.

Now we proceed to interpret the factor loadings obtained by modelling $\mathbf{R}$. It seems fairly clear from the factor loadings in Table 6 that the first factor, ${}_z\tilde{L}_1$, represents general economic conditions and might be called a market factor. All of the stocks load highly on this factor, and the loadings are about equal. The second factor contrasts the chemical stocks with the oil stocks. (The chemical companies have relatively large positive loadings, and the oil companies have large negative loadings, on the factor. The signs are reversed in comparison to what we obtained in the PC method but signs do not have any significance. The important observation is that they are different for the chemical and oil companies.) Thus, ${}_z\tilde{L}_2$ seems to differentiate stocks in different industries and might be called an industry factor. The first factor appears to be a market-wide effect on weekly stock price gains whereas the second factor reflects industry specific effects on chemical and oil stock price returns. To summarize, weekly stock price gains appear to be determined by general market conditions and activities that are unique to the different industries, as well as a residual or company specific factor.

Thus our interpretations of factors are exactly the same as before when estimation was performed using the PC method.

An idea of how well the estimated OFM model fits the covariance or correlation matrix when estimation is performed using Iterative PC method can be obtained by calculating the residual matrices $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathbf{T}} - \tilde{\mathbf{\Psi}}$ and $\mathbf{R} - {}_{\mathbf{z}}\tilde{\mathbf{L}}_{\mathbf{z}}\tilde{\mathbf{L}}^{\mathbf{T}} - {}_{\boldsymbol{z}}\tilde{\mathbf{\Psi}}$ (symbols represent obvious meanings) and observing whether the off-diagonal elements of the residual matrices are sufficiently close to zero.

$$\mathbf{S} - \left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T + \widetilde{\Psi}\right) = \begin{pmatrix} -0.00002356822 & 0.00003057901 & 0.00000399961 & -0.00002164349 & 0.00001679777 \\ 0.00003057901 & -0.00010763521 & 0.00001144688 & 0.00002442992 & -0.00001883313 \\ 0.00000399961 & 0.00001144688 & 0.00001260304 & 0.00000127725 & -0.00000229092 \\ -0.00002164349 & 0.00002442992 & 0.00000127725 & -0.00017026838 & 0.00006916144 \\ 0.00001679777 & -0.00001883313 & -0.00000229092 & 0.00006916144 & -0.00009966479 \end{pmatrix}$$

$$\mathbf{R} - \left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^T + {}_z\widetilde{\Psi}\right) = \begin{pmatrix} 0.000000000 & -0.000155164 & -0.001662220 & -0.021154225 & 0.031538322 \\ -0.000155164 & 0.000000000 & 0.001738787 & 0.018100937 & -0.027006673 \\ -0.001662220 & 0.001738787 & 0.000000000 & 0.002039266 & -0.003436274 \\ -0.021154225 & 0.018100937 & 0.002039266 & 0.000000000 & 0.002998902 \\ 0.031538322 & -0.027006673 & -0.003436274 & 0.002998902 & 0.000000000 \end{pmatrix}$$

The off-diagonal elements of both $\mathbf{S} - \left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^T + \widetilde{\Psi}\right)$ and $\mathbf{R} - \left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^T + {}_z\widetilde{\Psi}\right)$ are all very close to zero as expected. Thus there is an improvement in the estimates as compared to the previous estimates using PC method because there were some significant non-zero off-diagonal elements in $\mathbf{R} - \left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^T + {}_z\widetilde{\Psi}\right)$ previously. Since ease of factor interpretation is same in both methods, we consider the iterative PC estimates to be better than those obtained previously using PC method.

6.3 The Maximum Likelihood(ML) Method of estimation of the loading matrices

Suppose that we have n observations $\mathbf{x}_1, \ldots, \mathbf{x}_n$. If the common factors $\mathbf{F}$ and the specific factors $\boldsymbol{\varepsilon}$ can be assumed to be normally distributed, then maximum likelihood estimates of the factor loadings and specific variances may be obtained. Consider the case $\mathbf{F}_j$ and $\boldsymbol{\epsilon}_j$ are jointly normally distributed, the observations $\mathbf{x}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\epsilon}_j$ are then normal. Then, the likelihood is

$$\begin{aligned} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) =& (2\pi)^{-\frac{np}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2} \operatorname{tr}\left\{ \boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^{n} (\mathrm{x}_j - \overline{\mathrm{x}}) (\mathrm{x}_j - \overline{\mathrm{x}})^T + n(\overline{\mathrm{x}} - \boldsymbol{\mu})(\overline{\mathrm{x}} - \boldsymbol{\mu})^T \right) \right\} \right] \\ =& (2\pi)^{-\frac{(n-1)p}{2}} |\boldsymbol{\Sigma}|^{-\frac{(n-1)}{2}} \exp\left[-\frac{1}{2} \operatorname{tr}\left\{ \boldsymbol{\Sigma}^{-1} \left(\sum_{j=1}^{n} (\mathrm{x}_j - \overline{\mathrm{x}}) (\mathrm{x}_j - \overline{\mathrm{x}})^T \right) \right\} \right] \\ &+ (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{n}{2} (\overline{\mathrm{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\overline{\mathrm{x}} - \boldsymbol{\mu}) \right] \end{aligned} \tag{17}$$

and depends on L and Ψ through $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathbf{T}} + \boldsymbol{\Psi}$.

The $\mathbf{L}$ that maximizes the likelihood for this model is not unique, because of the multiplicity of the choices for L (by orthogonal transformation). To make L well defined, we impose a computationally convenient uniqueness condition $\mathbf{L}^T \Psi^{-1} \mathbf{L} = \Delta$, a diagonal matrix.

Maximizing the likelihood function with respect to (μ, L, Ψ) is not easy because the likelihood depends on the parameters in a very nonlinear fashion. Efficient numerical algorithm exists to maximize the likelihood iteratively and the MLEs can be obtained. The communalities for p variables with m factors and the proportion of the total variance accounted for by the j -th factor can be obtained similarly as in PC method.

6.3.1 Choice of the number of common factors m: A large sample test of goodness of fit

One of the main advantages of the Maximum Likelihood technique is that it provides a test of the hypothesis H_m that m common factors are sufficient to describe the data against the alternative that $\boldsymbol{\Sigma}$ has no constraints.

We want to test whether the m factor model appropriately describes the covariances among p variables i.e. we want to test

$$H_m : \boldsymbol{\Sigma}_{p\times p} = \mathbf{L}_{p\times m} \mathbf{L}^T_{m\times p} + \Psi_{p\times p} \text{ vs } H_1 : \boldsymbol{\Sigma} \text{ is any other p.d. matrix.}$$

An Likelihood Ratio (LR) test for H_m can be constructed as the ratio of the ML function under H_m and ML function under no restriction.. When $\boldsymbol{\Sigma}$ does not have any special form, the ML function is proportional to

$$|\mathbf{S}_n|^{-n/2} e^{-\frac{np}{2}}, \quad \mathbf{S}_n = \widehat{\boldsymbol{\Sigma}}_{H_1} = \frac{n-1}{n} \mathbf{S}$$

Under H_1, the multivariate normal likelihood is maximized at

$$\widehat{\mu} = \overline{\mathrm{x}}, \quad \widehat{\mathbf{\Sigma}}_{H_1} = \frac{1}{n}\sum_{j=1}^{n} (\mathrm{x}_j - \overline{\mathbf{x}}) (\mathbf{x}_j - \overline{\mathbf{x}})^T = \mathbf{S}_n$$

Under H_m, the multivariate normal likelihood is maximized at $\widehat{\mu} = \overline{\mathrm{x}}$, $\widehat{\mathbf{\Sigma}} = \widehat{\mathrm{L}}\widehat{\mathrm{L}}^T + \widehat{\Psi}$ where $\widehat{\mathrm{L}}$ and $\widehat{\Psi}$ are MLE of L and Ψ, respectively, under OFM with m common factors.

Then the ML function is proportional to

$$\begin{aligned} &|\widehat{\mathbf{\Sigma}}|^{-n/2} \exp\left(-\tfrac{1}{2}\operatorname{tr}\left\{\widehat{\mathbf{\Sigma}}^{-1}\left(\textstyle\sum_{j=1}^{n}(\mathbf{x}_j - \overline{\mathbf{x}})(\mathbf{x}_j - \overline{\mathbf{x}})^T\right)\right\}\right) \\ &= \left|\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\Psi}\right|^{-n/2} \exp\left(-\tfrac{n}{2}\operatorname{tr}\left\{\left(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\Psi}\right)^{-1}\mathbf{S}_n\right\}\right) \end{aligned}$$

The LR statistic for testing H_m is

$$\begin{aligned} -2\ln\Lambda &= -2\ln\left[\frac{\text{maximized likelihood under } H_m}{\text{maximized likelihood with no restriction}}\right] \\ &= -2\ln\left(\frac{\left|\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\Psi}\right|^{-n/2}}{|\mathbf{S}_n|^{-n/2}}\right) + n\operatorname{tr}\left[\left(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}}\right)^{-1}\mathbf{S}_n - p\right] \end{aligned}$$

with degrees of freedom

$$df = \frac{1}{2}p(p+1) - \left[p(m+1) - \frac{1}{2}m(m-1)\right] = \frac{1}{2}\left[(p-m)^2 - (p+m)\right]$$

Now $\operatorname{tr}\left(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n\right) - p = 0$ provided $\widehat{\Sigma} = \widehat{\mathrm{L}}\widehat{\mathrm{L}}^T + \widehat{\Psi}$ is the MLE of $\Sigma = \mathrm{L}\mathrm{L}^T + \Psi$.

Therefore, we have

$$-2\ln\Lambda = n\ln\left(\frac{|\widehat{\Sigma}|}{|\mathbf{S}_n|}\right) \tag{18}$$

Bartlett showed that χ^2 approximation to the sampling distribution is improved if n is replaced by

$$n' = (n-1) - \frac{1}{6}(2p+5) - \frac{2}{3}m$$

Then, the test statistic is

$$-2\ln\Lambda = n'\ln\left(\frac{\left|\widehat{\mathbf{L}}\hat{\mathbf{L}}^T + \widehat{\Psi}\right|}{|\mathbf{S}_n|}\right) \tag{19}$$

and reject H_m at level α if

$$-2\ln\Lambda > \chi^2_{df}$$

for large n and large $n-p$ where $df = \frac{1}{2}\left[(p-m)^2 - (p+m)\right]$.

To have $df > 0$, we must have

$$m < \frac{1}{2}[2p + 1 - \sqrt{8p+1}]$$

If data are not a random sample from multivariate normal distribution, this test tends to indicate the need for too many factors.

It can be shown that

$$\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|} = \frac{\left|\hat{\mathbf{L}}\hat{\mathbf{L}}^T + \hat{\Psi}\right|}{|\mathbf{S}_n|} = \frac{\left|\hat{\mathbf{L}}_z\hat{\mathbf{L}}_z^T + \hat{\Psi}_z\right|}{|\mathbf{R}|}$$

where the subscript z indicates standardized variables. Thus testing the validity of modelling the covariance matrix using m-factor model is equivalent to testing the validity of modelling the correlation matrix using m-factor model. So we need to perform this test only once either using the raw variables or the standardized variables.

6.3.2 Application of the Maximum Likelihood method to stock price data

The sample covariance matrix $\mathbf{S}$ and the sample correlation matrix $\mathbf{R}$ are as obtained in **Section 1** of this report. As in the application of the Principal Component(PC) and Iterative Principal Component(PC) method, we perform the factor analysis of both $\mathbf{S}$ and $\mathbf{R}$ employing $m = 2$ common factors. We later investigate the residual matrices and also perform a large sample test for checking whether using $m = 2$ common factors is appropriate or not.

Let $\widetilde{L} = \begin{bmatrix}\widetilde{L}_1 & \widetilde{L}_2\end{bmatrix}$ and ${}_z\widetilde{L} = \begin{bmatrix}{}_z\widetilde{L}_1 & {}_z\widetilde{L}_2\end{bmatrix}$ be the estimates of the loading matrices corresponding to the OFM imposed on $\mathbf{S}$ and $\mathbf{R}$ respectively.

The estimated factor loadings, communalities and specific variances using the Maximum Likelihood method are as follows:

Variable	Estimated Factor Loadings		Communalities	Specific variances
	$\tilde{L_1}$	$\tilde{L_2}$	$\tilde{h}_i^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$
X_1(Allied Chemical)	10^{-15}	10^{-15}	2×10^{-30}	0.0004332695
X_2(Du Pont)	10^{-15}	10^{-15}	2×10^{-30}	0.0004387172
X_3(Union Carbide)	10^{-15}	10^{-15}	2×10^{-30}	0.0002239722
X_4(Exxon)	10^{-15}	10^{-15}	2×10^{-30}	0.0007224964
X_5(Texaco)	10^{-15}	10^{-15}	2×10^{-30}	0.0007656742

Table 7: Estimates using Maximum Likelihood(ML) method for **S**

Variable	Estimated Factor Loadings		Communalities	Specific variances
	${}_z\tilde{L}_1$	${}_z\tilde{L}_2$	${}_z\tilde{h}_i^2$	${}_z\tilde{\psi}_i = 1 - {}_z\tilde{h}_i^2$
X_1(Allied Chemical)	0.1205972	0.754267066	0.5834625	0.41653750
X_2(Du Pont)	0.3284924	0.785749656	0.7253098	0.27469024
X_3(Union Carbide)	0.1876017	0.650216951	0.4579765	0.54202352
X_4(Exxon)	0.9974724	-0.007103504	0.9950016	0.00499844
X_5(Texaco)	0.6851746	0.026317440	0.4701568	0.52984315

Table 8: Estimates using Maximum Likelihood(ML) method for **R**

Comments/Interpretation of factors:

From the factor loadings corresponding to the covariance matrix **S** in Table 7, we see that all the variable loadings are essentially zero. So we conclude that interpretations based on these factor loadings are not useful.

Now we proceed to interpret the factor loadings obtained by modelling **R**. We observe that all variables have positive loadings on ${}_z\tilde{L}_1$. We call this factor the market factor representing general economic conditions, as we did in the principal component or iterative principal component solution. The oil company stocks load more heavily on this factor than the chemical company stocks. The interpretation of the second factor is not as clear as it appeared to be in the principal component solution. The chemical company stocks have large positive loadings and the oil company stocks have negligible loadings on the second factor ${}_z\tilde{L}_2$. From this perspective, the second factor differentiates the chemical company stocks from the oil company stocks and

might be called an industry factor. Alternatively, the second factor might be simply called a oil company factor.

The patterns of the initial factor loadings for the maximum likelihood solution are constrained by the uniqueness condition that $\hat{\mathbf{L}}'\hat{\Psi}^{-1}\hat{\mathbf{L}}$ be a diagonal matrix. Therefore, useful factor patterns may not be revealed until the factors are rotated, which we will study later.

An idea of how well the estimated OFM model fits the covariance or correlation matrix when estimation is performed using Maximum Likelihood method can be obtained by calculating the residual matrices $\mathbf{S} - \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathbf{T}} - \tilde{\mathbf{\Psi}}$ and $\mathbf{R} - {}_{\mathbf{z}}\tilde{\mathbf{L}}_{\mathbf{z}}\tilde{\mathbf{L}}^{\mathbf{T}} - {}_{\boldsymbol{z}}\tilde{\mathbf{\Psi}}$ (symbols represent obvious meanings) and observing whether the off-diagonal elements of the residual matrices are sufficiently close to zero.

$$\mathbf{S}-\left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{T}+\widetilde{\Psi}\right)=\begin{pmatrix}0.00043326946 & 0.0002756679 & 0.00015902652 & 0.00006411929 & 0.00008896616\\ 0.00027566790 & 0.0004387172 & 0.00017997369 & 0.00018145120 & 0.00012326228\\ 0.00015902652 & 0.0001799737 & 0.00022397219 & 0.00007341348 & 0.00006054612\\ 0.00006411929 & 0.0001814512 & 0.00007341348 & 0.00072249641 & 0.00050827720\\ 0.00008896616 & 0.0001232623 & 0.00006054612 & 0.00050827720 & 0.00076567417\end{pmatrix}$$

$$\mathbf{R}-\left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^{T}+{}_{z}\widetilde{\Psi}\right)=\begin{pmatrix}0.00000000000 & 0.00000748136 & -0.0025641467 & -0.0003325559 & 0.05198222326\\ 0.00000748136 & 0.00000000000 & 0.0016089457 & 0.0002116216 & -0.03307884578\\ -0.00256414670 & 0.00160894566 & 0.0000000000 & -0.0000095189 & 0.00055472164\\ -0.00033255593 & 0.00021162163 & -0.0000095189 & 0.0000000000 & 0.00012188721\\ 0.05198222326 & -0.03307884578 & 0.0005547217 & 0.0001218872 & 0.00000000000\end{pmatrix}$$

The off-diagonal elements of both $\mathbf{S}-\left(\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{T}+\widetilde{\Psi}\right)$ and $\mathbf{R}-\left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^{T}+{}_{z}\widetilde{\Psi}\right)$ are all very close to zero as expected. Thus there is an improvement in the estimates as compared to the previous estimates using PC method because there were some significant non-zero off-diagonal elements in $\mathbf{R}-\left({}_{\mathbf{z}}\widetilde{\mathbf{L}}_{\mathbf{z}}\widetilde{\mathbf{L}}^{T}+{}_{z}\widetilde{\Psi}\right)$ previously. However the estimates obtained using Iterative PC method provided us with better factor interpretations. We expect to obtain better interpretation using a rotated loading matrix as we will see later.

Since we have already verified that the assumption of multivariate normality of the data is plausible in **Section 3**, and both the number of observations $n = 103$ and $n - p = 103 - 5 = 98$ is large, we may use the large sample likelihood ratio test proposed by Bartlett to check the adequacy of the estimates obtained using Maximum Likelihood method assuming a 2-factor model. Following our discussion of this test of hypothesis previously, we obtain the observed value of the test statistic to be 2.00446333744836 and the degree of freedom is 1. Since the p-value of the observed value of the test statistic is 0.156836790051816 (or equivalently since the observed value of the test statistic is greater than the critical value $\chi^2_{0.05} = 3.84145882069412$), in the light of the given data, **we fail to reject the null hypothesis that a 2-factor model is adequate, at any reasonable level of significance**. **Thus our choice of $m = 2$ common factors is supported by the data.**

7 Varimax rotation of the loading matrices

As mentioned earlier, post-multiplying a matrix of factor loadings by any orthogonal matrix leads to the same approximation to the covariance (correlation) matrix. This means that mathematically, whether we estimate the loadings as $\widehat{\mathrm{L}}$ or as $\widehat{\mathrm{L}}^* = \widehat{\mathrm{L}}\mathrm{G}$, where G is orthogonal, the estimated covariance matrix remains unchanged.

$$\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}\mathbf{G}\mathbf{G}^T\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}} = \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*T} + \widehat{\mathbf{W}}$$

The estimated residual matrix also remain unchanged

$$\mathbf{S} - \widehat{\mathbf{L}}\widehat{\mathbf{L}}^T - \widehat{\Psi} = \mathbf{S} - \widehat{\mathbf{L}}^*\widehat{\mathbf{L}}^{*T} - \widehat{\mathbf{\Psi}}$$

The specific variances ψ_i and therefore the communalities h_i also remain unchanged. It is impossible to distinguish between loadings L and L^* from just a set of data even though in general they will be different.

Since only the loadings change by rotation, we rotate factors to see if we can better interpret the results. One possible objective to have each one of the p variables load highly on only one factor and have moderate to negligible loads on all other factors. It is not always possible to achieve this type of result. This can be examined with a varimax rotation, which we describe next.

Let $\widehat{\mathrm{L}}$ be the $p \times m$ matrix of estimated unrotated loadings and G be an $m \times m$ orthogonal matrix. The matrix of rotated loadings is

$$\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\mathbf{G}$$

Define $\tilde{l}^*_{ij} = \widehat{l}^*_{ij}/\widehat{h}_i$ as the scaled loading of the i -th variable on the j -th rotated factor, scaled by the square root of the communalities. We then compute the variance of the squares of the scaled loadings for the j -th rotated factor as follows

$$V_j = \frac{1}{p}\sum_{i=1}^{p}\left(\tilde{l}^*_{ij}\right)^2 - \frac{1}{p}\sum_{k=1}^{p}\left(\tilde{l}^*_{kj}\right)^2\Bigg)^2 = \frac{1}{p}\left(\sum_{i=1}^{p}\left(\tilde{l}^*_{ij}\right)^4 - \frac{1}{p}\left[\sum_{k=1}^{p}\left(\tilde{l}^*_{kj}\right)^2\right]^2\right)$$

The varimax procedure finds the orthogonal transformation of the loading matrix that maximizes the sum of those variances, summing across all m rotated factors, that is, $V = \sum_{j=1}^{m} V_j$. After rotation each of the p variables should load highly on at most one of the rotated factors.

Scaling the rotated coefficients $\hat{\ell}^*_{ij}$ has the effect of giving variables with small communalities relatively more weight in the determination of simple structure. After the transformation $\mathbf{T}$ is determined, the loadings $\tilde{\ell}^*_{ij}$ are multiplied by $\hat{h}_i$ so that the original communalities are preserved.

Note that $V \propto \sum_{j=1}^{m}($ Variance of squares of scaled loadings for j -th factor $)$

Maximizing V corresponds to 'spreading out' the squares of the loadings on each factor as much as possible. Therefore, we hope to find groups of large and negligible coefficients in any column of the rotated loadings matrix $\widehat{\mathbf{L}}^*$.

7.1 Application of varimax rotation to the estimates of the loading matrices in the stock price data model

Since the results obtained from the PC method and the iterative PC method were similar along with similar interpretations, we discuss the effects of varimax rotation only for the loading matrices obtained by modelling both the covariance and correlation matrices by application of the Iterative PC method and Maximum Likelihood method.

Let $\widetilde{L} = \begin{bmatrix} \widetilde{L}_1 & \widetilde{L}_2 \end{bmatrix}$ and ${}_z\widetilde{L} = \begin{bmatrix} {}_z\widetilde{L}_1 & {}_z\widetilde{L}_2 \end{bmatrix}$ be the estimates of the loading matrices corresponding to the OFM imposed on $\mathbf{S}$ and $\mathbf{R}$ respectively.

The estimated factor loadings, communalities and specific variances after applying varimax rotation using the Iterative PC and Maximum Likelihood method are as follows:

Variable	Estimated Factor Loadings		Communalities	Specific variances
	$\tilde{L}_1$	$\tilde{L}_2$	$\tilde{h}_i^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$
X_1(Allied Chemical)	0.001429288	0.015045796	0.0002284188	0.0002048506
X_2(Du Pont)	0.004713135	0.015841798	0.0002731762	0.0001655410
X_3(Union Carbide)	0.001651426	0.010146791	0.0001056846	0.0001182876
X_4(Exxon)	0.020796887	0.003724498	0.0004463824	0.0002761140
X_5(Texaco)	0.020606044	0.002839093	0.0004326695	0.0003330047

Table 9: Estimates using Iterative PC method with Varimax rotation for $\mathbf{S}$

Variable	Estimated Factor Loadings		Communalities	Specific variances
	${}_z\tilde{L}_1$	${}_z\tilde{L}_2$	${}_z\tilde{h}_i^2$	${}_z\tilde{\psi}_i = 1 - {}_z\tilde{h}_i^2$
X_1(Allied Chemical)	0.7567758	0.05291676	0.5755098	0.424490154
X_2(Du Pont)	0.8206362	0.21553342	0.7198985	0.280101509
X_3(Union Carbide)	0.6692376	0.10765378	0.4594683	0.540531664
X_4(Exxon)	0.1099817	0.99258935	0.9973296	0.002670401
X_5(Texaco)	0.1153959	0.67267227	0.4658042	0.534195813

Table 10: Estimates using Iterative PC method with Varimax rotation for $\mathbf{R}$

Variable	Estimated Factor Loadings		Communalities	Specific variances
	$\tilde{L_1}$	$\tilde{L_2}$	$\tilde{h}_i^2$	$\tilde{\psi}_i = s_{ii} - \tilde{h}_i^2$
X_1(Allied Chemical)	10^{-15}	10^{-15}	2^{-30}	0.0004332695
X_2(Du Pont)	10^{-15}	10^{-15}	2^{-30}	0.0004387172
X_3(Union Carbide)	10^{-15}	10^{-15}	2^{-30}	0.0002239722
X_4(Exxon)	10^{-15}	10^{-15}	2^{-30}	0.0007224964
X_5(Texaco)	10^{-15}	10^{-15}	2^{-30}	0.0007656742

Table 11: Estimates using Maximum Likelihood(ML) method with Varimax rotation for **S**

Variable	Estimated Factor Loadings		Communalities	Specific variances
	${}_z\tilde{L}_1$	${}_z\tilde{L}_2$	${}_z\tilde{h}_i^2$	${}_z\tilde{\psi}_i = 1 - {}_z\tilde{h}_i^2$
X_1(Allied Chemical)	0.7632891	0.02919252	0.5834625	0.41653750
X_2(Du Pont)	0.8194973	0.23180592	0.7253098	0.27469024
X_3(Union Carbide)	0.6680336	0.10820147	0.4579765	0.54202352
X_4(Exxon)	0.1126721	0.99111380	0.9950016	0.00499844
X_5(Texaco)	0.1083670	0.67706236	0.4701568	0.52984315

Table 12: Estimates using Maximum Likelihood(ML) method with Varimax rotation for **R**

We note that the estimated communalities and specific variances remain unchanged under varimax rotation as expected. Hence the residual matrices also remain unchanged.

Comments/Interpretation of factors:

From the factor loadings corresponding to the covariance matrix **S** in Tables 9 and 11, we see that all the variable loadings are essentially zero as before, which is expected. So we conclude that interpretations based on these factor loadings are not useful. We do observe though that, in the first factor, all the variables corresponding to the oil company stocks have factor loadings larger by an order of magnitude (i.e. about 10 times larger) in comparison to the factor loadings of the variables corresponding to the chemical company stocks. The converse statement holds true when we observe the loadings in the second factor where the variables corresponding to the chemical company stocks dominate the variables corresponding to the oil company stocks. This might indicate some structure in the covariance matrix which may be further revealed by the study of the correlation matrix **R**.

Now we proceed to interpret the factor loadings obtained by modelling **R**. Previously, when

varimax rotation was not applied, we had identified a "market" factor and an "industry" factor.

Irrespective of the method employed (Iterative PC or ML), the rotated loadings indicate that the variables corresponding to the chemical company stocks (Allied Chemical, Du Pont and Union Carbide) load highly on the first factor, while the variables corresponding to the oil company stocks (Exxon and Texaco) load highly on the second factor. The two rotated factors, together, differentiate the industries. It is difficult for us to label these factors intelligently. Factor 1 represents those unique economic forces that cause chemical company stocks to move together. Factor 2 appears to represent economic conditions affecting oil company stocks.

We note that the general factor (that is, one on which all the variables load highly) has been "destroyed" after varimax rotation is applied.

8 Estimation of the factor scores and verification of assumptions

In Factor Analysis, the interest is usually centered on the parameters present in the orthogonal factor model. However, estimated values of the common factors, called the factor scores are also required. These quantities are often used for diagnostic checking and as inputs to a subsequent study.

For $j = 1, \ldots, n$, the estimated m -dimensional vector of factor scores is

$$\widehat{\mathbf{f}}_j = \text{ estimate of the value } \mathbf{f}_j \text{ attained by } \mathbf{F}_j$$

Estimation situation is complicated by the fact that the unobserved quantities $\mathbf{f}_j$ and $\boldsymbol{\epsilon}_j$ outnumber the observed x_j. Our approach is to estimate $\mathbf{f}_j$, act as if factor loadings $\widehat{l}_{ij}$ and specific variances $\widehat{\psi}_i$ are true values. Usually, the estimated rotated loadings, rather than the original loadings are used to compute factor scores.

There are two different methods of computing the factor scores.

8.1 Weighted Least Squares Method

Suppose that the mean vector μ, the loading matrix L and specific variances Ψ are known for the model

$$\mathbf{X} - \mu = \mathbf{LF} + \boldsymbol{\epsilon}$$

We regard $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)^T$ as the error. Since $\text{Var}\,(\epsilon_i) = \psi_i, i = 1, \ldots, p,$, we use weighted least squares method to estimate the factor scores. This involves minimizing

$$\sum_{i=1}^{p} \frac{\epsilon_i^2}{\psi_i} = \boldsymbol{\epsilon}^T \boldsymbol{\Psi}^{-1} \boldsymbol{\epsilon} = (\mathbf{x} - \mu - \mathbf{Lf})^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \mu - \mathbf{Lf})$$

with respect to $\mathbf{f}$. Then

$$\widehat{\mathbf{f}} = \left(\mathbf{L}^T \boldsymbol{\Psi}^{-1} \mathbf{L}\right)^{-1} \mathbf{L}^T \boldsymbol{\Psi}^{-1} (\mathbf{x} - \mu)$$

Substituting the estimates of $\widehat{\mathbf{L}}, \widehat{\mathbf{\Psi}}$ and $\hat{\mu} = \overline{\mathbf{x}}$ in place of $\mathbf{L}, \mathbf{\Psi}$ and μ, we have the factor score of the j -th case as

$$\widehat{\mathbf{f}}_j = \left(\hat{\mathbf{L}}\widehat{\mathbf{\Psi}}^{-1}\hat{\mathbf{L}}^T\right)^{-1}\hat{\mathbf{L}}^T\hat{\mathbf{\Psi}}^{-1}\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)$$

If correlation matrix is used, that is, when standardized variables $\mathbf{z}_j, j = 1, \ldots, n$ are used.

$$\mathbf{z}_j = \mathbf{D}^{-\frac{1}{2}}\left(\mathbf{x}_j - \overline{\mathbf{x}}\right), \quad \mathbf{D} = \operatorname{diag}\left\{s_{11}, \ldots, s_{pp}\right\}$$

Then under orthogonal factor model $\boldsymbol{\rho} = \mathbf{L}_z\mathbf{L}_z^T + \mathbf{\Psi}_z$. Using the sample correlation matrix

$$\mathbf{R} = \widehat{\mathbf{L}}_z\widehat{\mathbf{L}}_z^T + \widehat{\mathbf{\Psi}}_z$$

the factor scores are

$$\widehat{\mathbf{f}}_j = \left(\widehat{\mathbf{L}}_z^T\widehat{\mathbf{\Psi}}_z^{-1}\hat{\mathbf{L}}_z\right)^{-1}\widehat{\mathbf{L}}_z^T\widehat{\mathbf{\Psi}}_z^{-1}\mathbf{z}_j, \quad j = 1, \ldots, n$$

If rotated loading $\widehat{\mathbf{L}}^* = \widehat{\mathbf{L}}\mathbf{G}, \mathbf{G}\mathbf{G}^T = \mathbf{I}$ is used, then $\widehat{\mathbf{f}}_j = \mathbf{G}^T\widehat{\mathbf{f}}_j, j = 1, \ldots, n$

8.2 Regression Method

In the regression method of estimation of factor scores, we start with the original factor model

$$\mathbf{X} - \mu = \mathbf{L}\mathbf{F} + \boldsymbol{\epsilon}$$

We assume that $\mathbf{L}$ and $\mathbf{\Psi}$ are known. If $\mathbf{F}$ and $\boldsymbol{\epsilon}$ are jointly normally distributed, then $(\mathbf{X} - \boldsymbol{\mu})$ has $\mathcal{N}_p\left(\mathbf{0}, \mathbf{L}\mathbf{L}^T + \mathbf{\Psi}\right)$ distribution and further,

$$\begin{pmatrix} \mathbf{X} - \boldsymbol{\mu} \\ \mathbf{F} \end{pmatrix} \sim \mathcal{N}_{p+m}\left(\mathbf{0}, \mathbf{\Sigma}^*\right), \quad \mathbf{\Sigma}^*_{(p+m)\times(p+m)} = \begin{pmatrix} \mathbf{L}\mathbf{L}^T + \mathbf{\Psi} & \mathbf{L} \\ \mathbf{L}^T & \mathbf{I} \end{pmatrix}$$

Therefore, the conditional distribution of $\mathbf{F}$ given $\mathbf{X} = \mathbf{x}$ is also normal with

$$E(\mathbf{F}|\mathbf{x}) = \mathbf{L}^T\mathbf{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \mathbf{L}^T\left(\mathbf{L}\mathbf{L}^T + \mathbf{\Psi}\right)^{-1}(\mathbf{x} - \boldsymbol{\mu})$$
$$\operatorname{Cov}(\mathbf{F}|\mathbf{x}) = \mathbf{I} - \mathbf{L}^T\mathbf{\Sigma}^{-1}\mathbf{L} = \mathbf{I} - \mathbf{L}^T\left(\mathbf{L}\mathbf{L}^T + \mathbf{\Psi}\right)^{-1}\mathbf{L}$$

Given any vector of observation $\mathbf{x}_j$ and taking the estimates $\hat{\mathbf{L}}$ and $\hat{\mathbf{\Psi}}$ as true values, the estimate of $\mathbf{f}_j$, the j -th factor score vector is given by

$$\widehat{\mathbf{f}}_j = \widehat{\mathbf{L}}^T\left(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\Psi}\right)^{-1}\left(\mathbf{x}_j - \overline{\mathbf{x}}\right)$$

Sometimes, to reduce the errors that may be included if the number of factors m is not quite appropriate, $\mathbf{S}$ is used in place of $\left(\widehat{\mathbf{L}}\widehat{\mathbf{L}}^T + \widehat{\mathbf{\Psi}}\right)$

8.2.1 Computation of estimates of factor scores and verification of assumptions for the stock price data

Since we obtain the best estimates and interpretations using the Iterative PC and Maximum Likelihood Methods in modelling the correlation matrix and the results given by the two methods are quite similar, we compute the factor scores and verify assumptions only for the rotated loadings corresponding to the modelling of the correlation matrix obtained using Maximum Likelihood Method.

The weighted least squares score estimates (first 4 and last 4 rows) are as follows:

$$\boldsymbol{f^{LS}} = \begin{pmatrix} 0.25089981 & -1.853673042 \\ 0.44303403 & 0.247877355 \\ -0.48078787 & -0.098356467 \\ 0.79921288 & -1.314979302 \\ \vdots & \vdots \\ 0.06316359 & -0.207170742 \\ 1.03125728 & -0.876430071 \\ 0.17139025 & -0.269180788 \\ -1.04440486 & -0.222903385 \end{pmatrix}$$

The regression score estimates (first 4 and last 4 rows) are as follows:

$$\boldsymbol{f^{R}} = \begin{pmatrix} 0.16535936 & -1.83427933 \\ 0.36753194 & 0.25550902 \\ -0.39519061 & -0.10792826 \\ 0.62520427 & -1.28789328 \\ \vdots & \vdots \\ 0.04719620 & -0.20424971 \\ 0.82432709 & -0.84768029 \\ 0.13434926 & -0.26346908 \\ -0.85866221 & -0.24362610 \end{pmatrix}$$

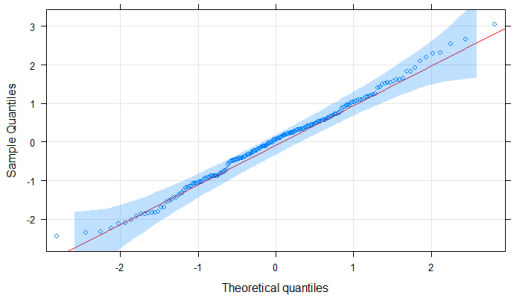
The first columns contain scores corresponding to the first factor while the second columns contain scores corresponding to the second factor.

Now we are going to check the assumption made on the factor scores, more specifically multivariate normality (here bivariate normality since $m = 2$ common factors are used) and pairwise independence (which, in case the scores corresponding to the 2 factors follow a bivariate normal distribution, reduces to checking whether they are uncorrelated).

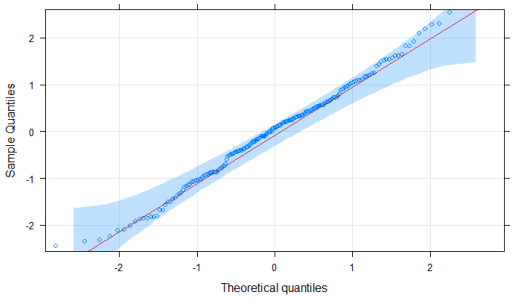
We first want to determine whether the factor scores corresponding to the two factors follows a bivariate normal distribution. We utilize the fact that the marginal distribution of any component of a random vector which follows bivariate normal is a univariate normal for a preliminary test of bivariate normality. For this we use

- **Graphical Method :-** We plot the theoretical quantiles from the univariate normal distribution against the sample quantiles and obtain the QQ plot for each of the factor scores corresponding to the two factors, the scores being computed using the weighted least squares and regression methods. Using the asymptotic distribution of sample quantiles from the normal distribution, a 95% confidence belt is also constructed for each of the QQ plots. The QQ plots are shown in Fig. 4.

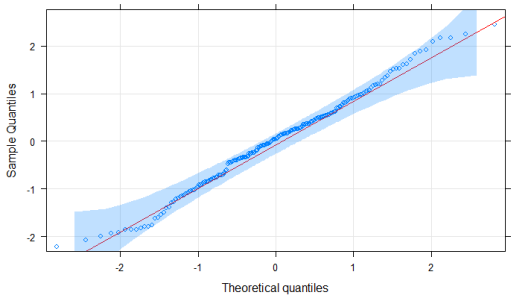
  Comments: From all the 4 QQ plots, we observe that there is appreciable agreement between sample and theoretical quantiles. Hence we see no significant visual evidence to reject the hypothesis that the marginal distributions of all the 4 factor scores are univariate normal.



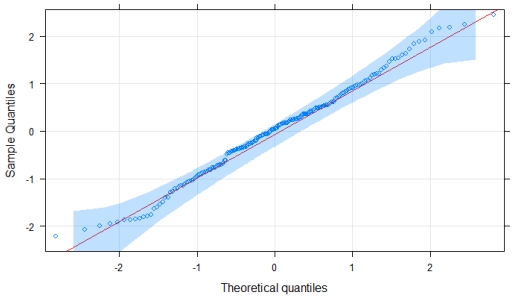
(a) QQ plot for scores corresponding to 1^{st} factor using WLS method



(b) QQ plot for scores corresponding to 2^{nd} factor using WLS method



(c) QQ plot for scores corresponding to 1^{st} factor using Regression method



(d) QQ plot for scores corresponding to 2^{nd} factor using Regression method

Figure 4: QQ plots of factor scores

- **Formal testing procedure :-** To validate our observations from the QQ plots, we perform the Shapiro-Wilks Test for univariate normality for each of the 4 factor scores.

  The p-values of the Shapiro-Wilks tests are summarized in the following table:

Variable	p-value
WLS Scores for 1^{st} factor	0.7779096
WLS Scores for 2^{nd} factor	0.3497038
Regression Scores for 1^{st} factor	0.7423269
Regression Scores for 2^{nd} factor	0.3830935

  At any reasonable level of significance we can therefore conclude that the marginal distributions of the variables are univariate normal.

  However multivariate normality is not ensured even if the marginal distributions are univariate normal. To check whether the corresponding follows a multivariate (here bivariate) normal distribution, we use the multivariate extension of the Shapiro Wilks test of normality proposed by Villasenor-Alva and Gonzalez-Estrada [2], available in the **mvShapiroTest** package in R. The p-values of the tests are obtained as follows.

Variable	p-value
WLS Scores	0.7779096
Regression Scores	0.7423269

  Hence the null hypothesis of multivariate (here bivariate) normality of the factor scores using the two methods is accepted at any reasonable level of significance.

Since the factor scores follow bivariate normal distributions, we can test for pairwise independence of the factor scores corresponding to the two factors using Pearson's test of independence based on the correlation coefficient between the factor scores corresponding to the two factors. Under the null hypothesis H_0 it is assumed that the two factor scores follow univariate normal distributions which are independent of each other. When bivariate normality holds, this is equivalent to $H_0 : \rho = 0$ where ρ is the population correlation coefficient between the factor scores. We test H_0 against the alternative hypothesis $H_1 : \rho \neq 0$ using the test statistic $t = r\sqrt{\frac{n-2}{1-r^2}}$ where n is the length of the factor scores vector and r is the Pearson's product moment correlation coefficient between the factor scores corresponding to the two factors. under H_0, the test statistic has a Student's t distribution with $n-2$ degrees of freedom. We reject H_0 against H_1 at α level of significance if $|t_{obs}| > t_{n-2,\frac{\alpha}{2}}$ where $t_{n-2,\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ point of the Student's t distribution with $n-2$ degrees of freedom.

The p-values of the two correlation tests are obtained as follows:

Variable	p-value
WLS Scores	0.8112496
Regression Scores	0.8112538

Hence, in the light of the given data, the null hypothesis that the factor scores are pairwise independent in both cases is accepted at any reasonable level of significance.
Thus we have verified that the assumptions that the factor scores follow a multivariate normal distribution and are pairwise independent are plausible in the light of the given data.

9 References

[1] A. Wald and J. Wolfowitz. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162, 1940.

[2] José A. Villasenor Alva and Elizabeth González Estrada. A generalization of shapirowilk's test for multivariate normality. *Communications in Statistics - Theory and Methods*, 38(11):1870–1883, 2009.

[3] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007.