Multivariate Models – Elliptical Distribution

Owen Adhikaputra, Longhao Jin, Yice Luo, Linfeng Zhang

December 3, 2018

Contents

Part I

Theoretical Introduction

1 Basic of Multivariate Modeling

1.1 Random Vectors and Their Distribution

Joint and marginal distribution

Given a general d-dim random vector $X = (X_1, ..., X_d)'$, we have cdf

$$F_{\mathbf{X}}(\mathbf{x}) = P(X_1 \leqslant x_1, ..., X_d \leqslant x_d). \tag{1.1}$$

For all i in {1,. . . ,d}, we have marginal df

$$F_i(x_i) = P(X_i \leqslant x_i) = F(\infty, ..., x_i, ...\infty). \tag{1.2}$$

Equation (1.2) can be extend to k-dim case, for $2 \leqslant k \leqslant d$. Partition $\mathbf{X}$ into $(\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1 = (X_1, ..., X_k)'$, then the marginal df of $\mathbf{X}_1$ is

$$F_{\mathbf{X}_1}(\mathbf{x}_1) = P(\mathbf{X}_i \leqslant \mathbf{x}_i) = F(x_1, ..., x_k, \infty, ...\infty). \tag{1.3}$$

The df of a random vector X is said to be absolutely continuous if

$$F(x_1, ..., x_d) = \int_{-\infty}^{x_1} ... \int_{-\infty}^{x_d} f(u_1, ..., u_d) du_1 ... du_d \tag{1.4}$$

for some non-negative function f, which is then known as the joint density of X.

Note: the existence of a joint density implies the existence of marginal densities for all k-dimensional marginals. However, the existence of a joint density is not necessarily implied by the existence of marginal densities.

Conditional distributions and independence

Partition $\mathbf{X}$ into $(\mathbf{X}_1^t, \mathbf{X}_2^t)^t$, where $\mathbf{X}_1 = (X_1, ..., X_k)^t$, then the conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1 = \mathbf{x}_1$ has density

$$f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) = \frac{f(\mathbf{X}_2, \mathbf{X}_1)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} = F(x_1, ..., x_d) = \int_{u_{k+1}=-\infty}^{x_{k+1}} ... \int_{u_d=-\infty}^{x_d} f(u_1, ..., u_d) du_1 ... du_d. \tag{1.5}$$

If and only if $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent, then

$$F(\mathbf{x}) = F_{\mathbf{X}_1}(\mathbf{x}_1) F_{\mathbf{X}_2}(\mathbf{x}_2), \quad \forall \mathbf{x} \in R^d \tag{1.6}$$

In this case where $\mathbf{X}$ possesses a density, $f(\mathbf{x}) = f_{\mathbf{X}_1}(\mathbf{x}_1) f_{\mathbf{X}_2}(\mathbf{x}_2)$. The components of $\mathbf{X}$ are mutually independent if and only if $F(\mathbf{x}) = \Pi_{i=1}^d F_i(x_i), \ \ \forall x \in R^d$.

Moments and characteristic function

The mean vector of $\mathbf{X}$ is given by $E(\mathbf{X}) = (E(X_1), ..., E(X_d))'$ if it exists. The covariance matrix is given by $cov(\mathbf{X}) = E((\mathbf{X} - E(\mathbf{X})(\mathbf{X} - E(\mathbf{X})')$, often written as Σ. The (i,j)th element of Σ is

$$\sigma_{ij} = cov(X_i, X_j) = E(X_i X_j) - E(X_i) E(X_j) \tag{1.7}$$

The correlation matrix of $\mathbf{X}$ is denoted by $\rho(\mathbf{X})$. The $(i,j)^{th}$ element of $\rho(\mathbf{X})$ is

$$\rho_{ij} = \rho(X_i, X_j) = \frac{cov(X_i, X_j)}{\sqrt{var(X_i) var(X_j)}} \tag{1.8}$$

Introduce two operators on a covariance matrix Σ:

$$\Delta(\Sigma) = diag(\sqrt{\sigma_{11}}, ..., \sqrt{\sigma_{dd}}) \tag{1.9}$$

$$\wp(\Sigma) = (\Delta(\Sigma))^{-1} \Sigma (\Delta(\Sigma))^{-1} \tag{1.10}$$

Then we have

$$\rho(\mathbf{X}) = \wp(\Sigma) \tag{1.11}$$

For any matrix $\mathbf{B} \in \mathbf{R}^{k \times d}$ and vector $\mathbf{b} \in \mathbf{R}^k$, we have

$$E(\mathbf{BX} + \mathbf{b}) = \mathbf{B}E(\mathbf{X}) + \mathbf{b} \tag{1.12}$$

$$cov(\mathbf{BX} + \mathbf{b}) = \mathbf{B}cov(\mathbf{X})\mathbf{B}' \tag{1.13}$$

Note that equation (1.12) implies that

$$var(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a} \geqslant 0 \qquad \forall \mathbf{a} \in \mathbf{R}^d$$

and Σ is positive semidefinite. When

$$\mathbf{a}'\Sigma\mathbf{a} > 0 \qquad \forall \mathbf{a} \in \mathbf{R}^d \backslash \{\mathbf{0}\},$$

Σ is positive definite and invertible.

Using Cholesky factorization, the covariance matrix can be written as $\Sigma = AA'$ for a lower triangular matrix A with positive diagonal elements. The matrix A is known as the Cholesky factor, denoted by $\Sigma^{1/2}$ for convenience.

1.2 Standard Estimators of Covariance and Correlation

Assumptions

In this report, we assume that the observations are identically distributed or at least serially uncorrelated. (Serial correlation is the relationship between a given variable and a lagged version of itself over various time intervals.)

For shorter time intervals independence may be a less appropriate assumption (due to a phenomenon known as volatility clustering), but the assumption of independence may be roughly tenable for longer time intervals such as months or years and serial correlation of returns is often quite weak.

Estimators

Standard method-of-moments estimators of μ and Σ are given by the sample mean vector $\bar{\mathbf{X}}$ and the sample covariance matrix S. They are defined by

$$\bar{\mathbf{X}} = \frac{1}{n}\Sigma_{i=1}^{n}\mathbf{X}_i, \qquad \mathbf{S} = \frac{1}{n}\Sigma_{i=1}^{n}(\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})', \tag{1.14}$$

where $\bar{\mathbf{X}}$ is unbised but S is biased. $(n(\mathbf{S})/(n-1)$ is an unbiased estimator for Σ.)

The $(j,k)^{th}$ element of The sample correlation matrix R is given by $r_{jk} = s_{jk}/\sqrt{s_{jj}s_{kk}}$. Or by (1.11), we have

$$\mathbf{R} = \wp(\mathbf{S})$$

.

Discussion

If our data $X_1, \ldots, X_n$ are iid multivariate normal, then $\bar{X}$ and S are the maximum likelihood estimators (MLEs) of the mean vector μ and covariance matrix Σ. Their behavior as estimators is well understood while the multivariate normal is certainly not a good description of financial risk-factor returns over short time intervals.

Under these circumstances the behavior of the standard estimators in (1.14) is often less well understood, and other estimators of the true mean vector μ and covariance matrix Σ may perform better in terms of efficiency and robustness. (A more efficient estimator is an estimator with a smaller expected estimation error; a more robust estimator is an estimator who is not so influenced by outliers.)

1.3 The Multivariate Normal Distribution

Definition 1.3.1. *$X = (X_1, ..., X_d)'$ has a multivariate normal or Gaussian distribution if*

$$\boldsymbol{X} \stackrel{d}{=} \mu + \boldsymbol{A}\boldsymbol{Z},$$

where $\boldsymbol{Z} = (Z_1, ..., Z_k)'$ is a vector of iid univariate standard normal rvs (mean 0 and variance 1), and $\boldsymbol{A} \in \boldsymbol{R}^{d\times k}$ and $\mu \in \boldsymbol{R}^d$ are respectively a matrix and a vector of constants.

The characteristic function of a standard univariate normal variate Z is $\phi_Z(t) = e^{-t^2/2}$. Therefore, the characteristic function of $\mathbf{X}$ can be calculated to be

$$\phi_X(\mathbf{t}) = E(e^{it'X}) = exp(it'\mu - 1/2t'\Sigma t), \qquad \mathbf{t} \in \mathbf{R}^d \tag{1.15}$$

We concentrate on the non-singular case of the multivariate normal when $rank(A) = d \leqslant k$. In this case the covariance matrix Σ has full rank d and is therefore invertible (non-singular) and positive definite. Moreover, X has an absolutely continuous distribution function with joint density given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp\{-1/2(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\}, \qquad \mathbf{x} \in \mathbf{R}^d \tag{1.16}$$

We can see from (1.16) that points with equal density lie on ellipsoids determined by equations of the form

$$(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) = c, \quad \forall c > 0. \tag{1.17}$$

Whenever a multivariate density $f(x)$ depends on x only through the quadratic form $(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)$, the distribution is called **elliptical distribution**.

Algorithm 1.3.1. *simulation of multivariate normal distribution.*

(1) Perform a Cholesky decomposition of Σ to obtain the Cholesky factor $\Sigma^{1/2}$

(2) Generate a vector $\boldsymbol{Z} = (Z_1, ..., Z_d)'$ *of independent standard normal variables.*

(3) Set $\boldsymbol{X} = \mu + \Sigma^{1/2}\boldsymbol{Z}$

The L.H.S. of the figure 1.1 represents simulated data and the R.H.S. represents the underlying true distribution. Comparing L.H.S. and R.H.S, the simulation looks good. Each point on the same contour has the same probability, which is given by the number on the contour. Correlation=0.7 gives an obvious trend that x_2 increases in probability as x_1 increases; correlation = -0.7 shows the opposite; when correlation = 0.3, the trend is still observable but not as obvious.

Properties of the Multivariate Normal

These properties underline the attractiveness of the multivariate normal for computational work in risk management. Many of them are in fact shared by the broader classes of normal mixture distributions and elliptical distributions.

1. Linear Combination Using the characteristic function (1.15), we can show that the linear combinations of multivariate normal random vectors remain multivariate normal,and

$$\mathbf{BX}+\mathbf{b} \sim N_k(\mathbf{B}\mu + \mathbf{b}, \mathbf{B\Sigma B}'), \forall \mathbf{B} \in \mathrm{R}^{k\times d}, \mathbf{b} \in \mathbf{R}^k, \tag{1.18}$$

2. Marginal Distributions Partition $\mathbf{X}$ into $(\mathbf{X}_1', \mathbf{X}_2')'$, where $\mathbf{X}_1 = (X_1, ..., X_k)'$. Similarly, we can extend the notation to μ and Σ. We have $\mathbf{X}_1 \sim N_k(\mu_1, \Sigma_{11})$ and $\mathbf{X}_2 \sim N_k(\mu_2, \Sigma_{22})$.

3. Conditional Distributions Assuming that Σ is positive definite, the conditional distributions of X_2 given X_1 and of X_1 given X_2 may also be shown to be multivariate normal. For example, $\mathbf{X_2}|\mathbf{X_1} = \mathbf{x_1} \sim N_{d-k}(\mu_{2.1}, \Sigma_{22.1})$, where

$$\mu_{\mathbf{2.1}} = \mu_{\mathbf{2}} + \mathbf{\Sigma_{21}\Sigma_{11}^{-1}}(\mathbf{x_1} - \mu_{\mathbf{1}})$$

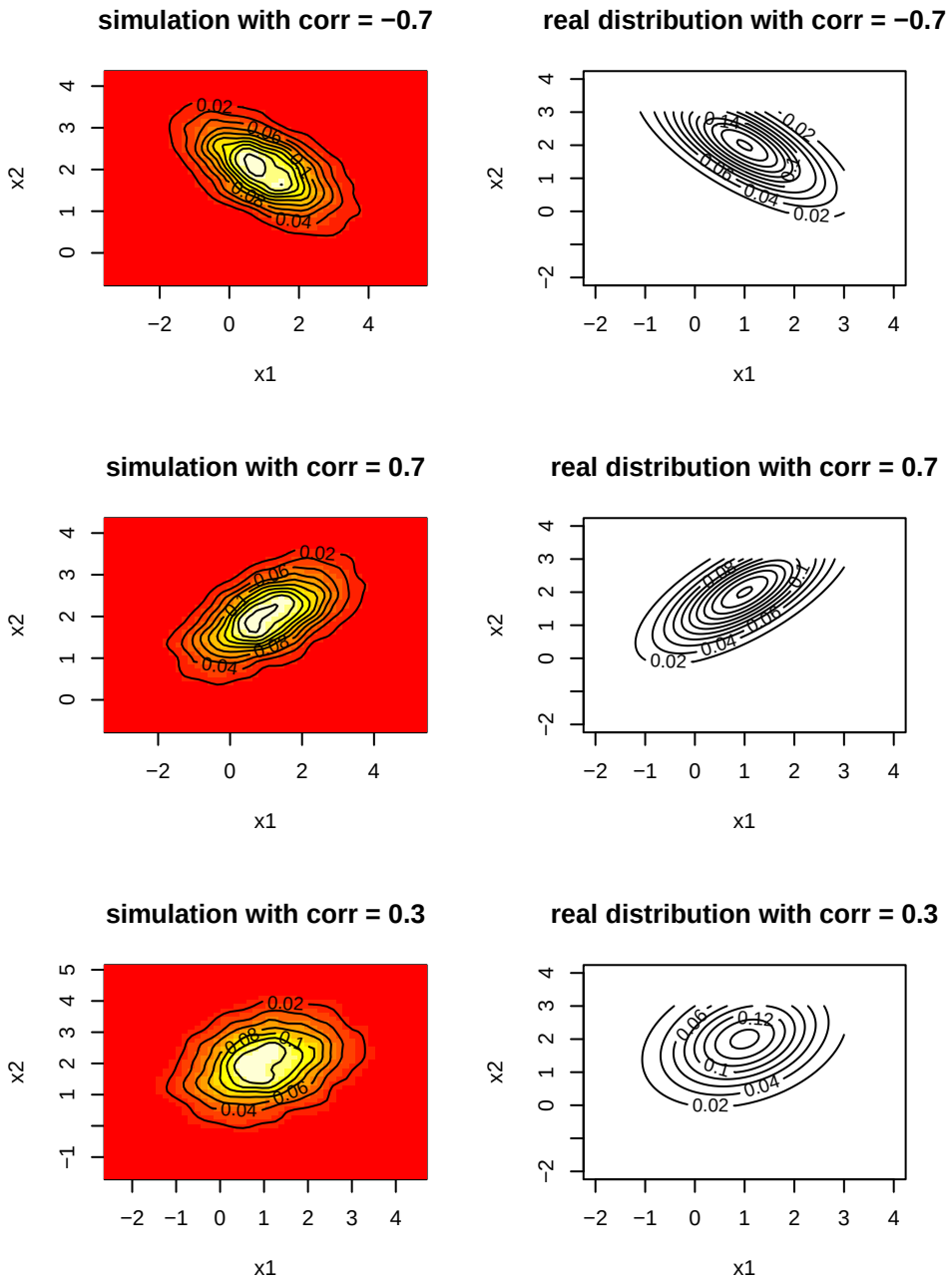

Figure 1.1: simulation of multivariate normal distribution

and

$$\mathbf{\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}}$$

are the conditional mean vector and covariance matrix.

4. Quadratic Form if $\mathbf{X} \sim N_d(\mu, \Sigma)$ with positive definite Σ, then $\mathbf{(X-\mu)\Sigma^{-1}(X-\mu) \sim \chi_d^2}$. This property is used for checking normality.

5. Convolutions if $\mathbf{X} \sim N_d(\mu, \mathbf{\Sigma})$ and $\mathbf{Y} \sim N_d(\tilde{\mu}, \tilde{\mathbf{\Sigma}})$ are independent,then we can show that $\mathbf{X+Y} \sim N_d(\mu+\tilde{\mu}, \mathbf{\Sigma}+\tilde{\mathbf{\Sigma}})$

1.4 Testing Multivariate Normality

Univariate tests

If $\mathbf{X_1}, ..., \mathbf{X_n}$ are iid multivariate normal, then for $1 \leq j \leq d$ the univariate sample $X_{1,j}, \ldots, X_{n,j}$ consisting of the observations of the jth component must be iid univariate normal; in fact, any univariate sample constructed from a linear combination of the data of the form $\mathbf{a'X_1}, ..., \mathbf{a'X_n}$ must be iid univariate normal. This can be assessed graphically with a Q-Q plot against a standard normal reference distribution.

multivariate tests

To test that the univariate margins of the distribution are normal is not sufficient to test for multivariate normality. Thus we also need to be able to test joint normality, and a simple way of doing this is to exploit the fact that the quadratic form in (1.17) has a chi-squared distribution.

$$\{D_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}) : i = 1, ..., n\}. \tag{1.19}$$

D_i^2s are not under exactly chi-squared distribution because $X_i - \bar{\mathbf{X}}$ and S are not independent. However, we construct QQ plots against D_i^2s distribution to test the normality,

because it's close to chi-squared for large n. (The true distribution is

$$n(n-1)^{-2}D_i^2 \sim Beta(d/2,(n-d-1)/2).$$

Define skewness and kurtosis to be

$$\mathbf{b}_d = 1/n^2\Sigma_{i=1}^n\Sigma_{j=1}^n D_{ij}^3, \quad k_d = 1/n\Sigma_{i=1}^n D_i^4, \tag{1.20}$$

where D_i is given by (1.19) known as Mahalanobis distance and $D_{ij}(\mathbf{X}_i-\bar{\mathbf{X}})'\mathbf{S}^{-1}(\mathbf{X}_j-\bar{\mathbf{X}})$. Under null assumption of multivariate normality, the asymptotic distributions of these statistics as n goes to infinity are

$$\frac{1}{6}n\mathbf{b}_d \sim \chi^2_{d(d+1)(d+2)}, \qquad \frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim N(1,0). \tag{1.21}$$

Mardia's test of multinormality involves comparing the skewness and kurtosis statistics with the above theoretical reference distributions.

2 Normal Mixture Distributions

This section will be a direct continuation from the previous part as we will need most of the concepts we learned before. Normal mixture distributions are the more advanced form of multivariate normal distribution since we introduced randomness into first the covariance matrix and also the mean vector of a multivariate normal distribution. The term that we will use for this is W. First,we will take a look at one of the forms of normal mixture distributions, which is the normal variance mixture distribution.

2.1 Normal Variance Mixture Distribution

Definition 2.1.1. *The random vector X is said to have a (multivariate) normal variance mixture distribution if*

$$X \overset{d}{=} \mu + \sqrt{W}\ AZ$$

where,

(i) $Z \sim N_k\ (0,\ I_k\ )$*;*

(ii) $W \geq 0$ *is a non-negative, scalar-valued rv that is independent of* Z*, and*

(iii) $A \in R^{d \times k}$ *and* $\mu\ R^d$ *are a matrix and a vector of constants, respectively.*

Now we have seen the definition for normal variance mixtures, such that we can see the connection to multivariate normal. Recall that multivariate normal has a normal distribution of $X \overset{d}{=} N_d\ (\mu,\ \Sigma)$ and compare it to normal variance mixture: $X \mid W \overset{d}{=} \sim w\ N_d\ (\mu,\ w\ \Sigma)$, where $\Sigma = AA'$. This is not a coincidence as we can think that the distribution of X as a combination of multivariate normal distributions with the same mean vector and the same covariance matrix, but with multiplication to a constant w describe before. In other words, the construction of normal mixture distribution is based on drawing randomly from this set of components multivariate normals according to a set of weights, or we call it W; where the result is no longer a multivariate normal distribution. Thus, W can be described as a shock from new information we get and such impacting volatility in risk factors.

Following the multivariate normal, we are only interested in normal mixture distribution if $\text{rank}(A) = d \leq k$ and Σ is a full-rank, positive definite matrix, such that it will give us a non-singular normal variance mixture or it has a probability density function.

Keeping W has a finite expectation, we can calculate both expected and covariance of X

$$E\ (X) = E\ (\mu + \sqrt{W}\ AZ) = \mu + E(\ \sqrt{W})\ A\ E\ (Z) = \mu$$

also

$$cov\ (X) = E\ ((\sqrt{W}\ AZ)(\sqrt{W}\ AZ)') = E(\ W)\ A\ E\ (ZZ')\ A' = E\ (W)\ \Sigma$$

Moreover, we can calculate the characteristic function of a normal variance:

$$\Phi_x\ (t) = E(E(e^{it'X}\ |\ W)) = E(exp(it'\mu - \tfrac{1}{2}\ Wt'\ \Sigma\ t) = e^{it'\mu}\ \hat{H}\ (\tfrac{1}{2}\ t'\ \Sigma\ t)$$

where $\hat{H}\ (\theta) = \int_0^\infty e^{-\theta v} dH(v)$ is the Laplace-Stieltjes transform of the distribution function H of W. Hence, we can use the notation $X \sim M_d\ (\mu,\ \Sigma,\ \hat{H})$ for normal variance mixtures.

The last point that we want to look at is the conditional density of X given W. To get it; we need to assume that Σ is positive definite and the distribution of W has no point mass at zero, such that the density of X is given by

$$f(x) = \int\ \frac{w^{-d/2}}{(2\pi)^{d/2}|\Sigma|^{1/2}}\ exp\ \{-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2w}\}\ dH(w)$$

Notice the form $-(x-\mu)'\Sigma^{-1}(x-\mu)$ which mean they are the densities of elliptical distributions.

Example 2.1.1. ***Multivariate two-point normal mixture distribution***

Simple examples of normal mixtures are obtained when W is a discrete rv. For example, the two-point normal mixture model is obtained by taking W to be a discrete random variable that assumes the distinct positive values $k1$ and $k2$ with probabilities p and $1-p$, respectively. By setting$k2$ large relative to $k1$ and choosing p large, this distribution might be used to define two regimes: an ordinary regime that holds most of the time and a stress regime that occurs with small probability $1-p$. Obviously this idea extends to k-point mixture models.

Example 2.1.2. ***Multivariate t distribution***

In this example, we are going to further define W as a random variable with an inverse gamma distribution $W \sim Ig(\frac{1}{2}v, \frac{1}{2}v)$, which means now X has a multivariate t distribution with v degrees of freedom. Now we use $X \sim t_d\ (v, \mu, \Sigma)$ as multivariate t distribution

notation, where Σ in here is not the covariance matrix of X. We can see the reason if we look at our new expected value of $E(W)$ and $cov(X)$, where $E(W) = \frac{v}{(v-2)}$ (note that mean of inverse gamma is $\frac{\beta}{\alpha-1}$) and $cov(X) = (\frac{v}{(v-2)})\Sigma$, such that the covariance is only defined if $v > 2$.

Using the pdf of normal mixture distributions we can get the density of t distribution

$$f(x) = \frac{\Gamma(\frac{1}{2}(v+d))}{\Gamma(\frac{1}{2}(v)(\pi v)^{d/2}|\Sigma|^{1/2}} \ (1 + \frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{v})^{-(v+d)/2}$$

We can see that this is another clear evidence that multivariate t distribution is an ellipsoid with equation $-(x-\mu)'\Sigma^{-1}(x-\mu) = c$, where $c > 0$. We also have a nice figure to compare between multivariate t distribution with multivariate normal. From one quick look, everyone can determine that multivariate t distribution has denser countour in the center of graph compare to multivariate normal. Multivariate t distribution also has heavier marginal tails (events that are much larger than the mean happen frequently) and a more pronounced tendency to generate simultaneous extreme (where in real life case we can see 80% of the wealth owned by 20% of the population) compared to multivariate normal.

Example 2.1.3. ***Symmetric generalized hyperbolic distribution***

The last example we will encounter is a flexible family of normal variance mixtures with taking W to be a random variable with a generalized inverse Gaussian (GIG) distribution, such that $W \sim N^{-1}$ (λ, χ, ψ).

We will see the more general case of this distribution later throughout this paper using mean-variance mixtures of normal, and it is not necessarily elliptical distribution.

Linear Combinations of Normal Variance Mixtures

$$\text{If } X \sim M_d(\mu, \Sigma, \hat{H}) and \ Y = Bx + b, where B \in R^{d\times k} \ and$$

$$b \in R^d, then Y \sim M_d(B\mu + b, B\Sigma B', \hat{H}).$$

We can proof this by the same manner as before

$$\Phi_Y\ (t) = E(e^{it'(BX+b)}) = e^{it'b}\Phi_X(B't) = e^{it'(B\mu+b}\ \hat{H}\ (\tfrac{1}{2}\ t'\ B\Sigma B'\ t)$$

The subclass of mixture distributions specified by $\hat{H}$ is therefore closed under linear transformations. This means hat if X has a multivariate t distribution with v degress of freedom, then so does any linear transformation of X

Algorithm 2.1.4. (Simulate a Normal Variance Mixture).

Identical step with multivariate normal distribution can be taken to simulate normal variance mixture

(1) Generate $Z \sim N_d\ (0, \Sigma\ )$ using $e^{it'\mu}\ \hat{H}\ (\frac{1}{2}\ t'\ \Sigma\ t)$.

(2) Generate independently a positive mixing variable W with distribution function H (corresponding to the Laplace–Stieltjes transform $\hat{H}$).

(3) Set $X \stackrel{d}{=} \mu + \sqrt{W}\ AZ$

2.2 Normal Mean-Variance Mixtures

All of the examples that we tried before are not suitable to use for analyzing a real risk-factor return data since it is oversimplified and also can be classified as an elliptical symmetry. Properties of elliptical symmetry imply that all one-dimensional marginal distributions are rigidly symmetric, which mean it has the same frequent number of negative return with a positive return of a stock. Obviously, this is not the case in a real-life situation where negative returns (losses) have heavier tails (higher probability) compare to positive returns (gains). These problems can be tackled by mixing normal distributions with different means as well as different variances such that the new mixture

has an asymmetry on it. We will call this new mixture as a normal mean-variance mixtures.

Definition 2.2.1. *The random vector X is aid to have a (multivariate) normal mean-variance mixture distribution if*

$$X \stackrel{d}{=} w(W) + \sqrt{W}\, AZ$$

where,

(i) $Z \sim N_k\ (0,\ I_k\ )$*;*

(ii) $W \geq 0$ *is a non-negative, scalar-valued random variable that is independent of* Z*,*

(iii) $A \in R^{d\times k}$ *is a matrix and,*

(iv) $m : [0, \infty) \to R^d$ *is a measurable function*

We can use the same form as normal variance mixtures to get the equation for normal mean-variance mixture:

$$X \mid W \stackrel{d}{=} w \sim N_d\ (m(v),\ w\ \Sigma), \text{ where } \Sigma = AA'$$

A possible concrete specification for the function $m(W)$ is

$$m(W) = \mu + W\gamma$$

where μ and γ are parameter vectors in R^d. Since we also know that $E(X|W) = \mu + W\gamma$ and $cov(X|W) = W\Sigma$, we can find the expected and covariance of X

$$E(X) = E(E(X|W) = \mu + e(W)\gamma$$

$$cov(X) = E(cov(X|W) + cov(E((X|W)) = E(W)\Sigma + var(W)\gamma\gamma'$$

when the mixing variable W has finite variance. Notice that the expected and covariance of X is the same normal variance mixtures if $\gamma = 0$ since we do not introduce any skewness into our mixture.

2.3 Generalized Hyperbolic Distributions

In Example 2.1.3 (Symmetric generalized hyperbolic distribution), we saw the special case of generalized hyperbolic (GH) distributions consisting of the elliptically symmetric normal variance mixture distributions. We can obtain the GH by using both$X \stackrel{d}{=} w(W) + \sqrt{W}\ AZ$ and $X \mid W \stackrel{d}{=} w \sim N_d\ (m(v),\ w\ \Sigma)$. We are also assuming that $W \sim N^{-1}\ (\lambda,\ \chi,\ \psi)$, a GIG (generalized inverse Gaussian) distribution.

We can see the joint density for GH is

$$f(x) = \int_0^\infty \frac{e^{(x-\mu)'\Sigma^{-1}\gamma}}{(2\pi)^{d/2}|\Sigma|^{1/2}w^{1/2}} \times exp\{-\frac{(x-\mu)'\Sigma^{-1}(x-\mu)}{2w} - \frac{\gamma'\Sigma^{-1}\gamma}{2/w}\}h(w)dw$$

where $h(w)$ is the density of W. Evaluation of this integral gives the GH density

$$f(x) = c\frac{K_\lambda - (d/2)\sqrt{(\chi+(x-\mu)'\Sigma^{-1}(x-\mu))(\psi+\gamma'\Sigma^{-1}\gamma))}e^{(x-\mu)'\Sigma^{-1}\gamma}}{\sqrt{(\chi+(x-\mu)'\Sigma^{-1}(x-\mu))(\psi+\gamma'\Sigma^{-1}\gamma))}^{(d/2)-\lambda}}$$

where the normalizing constant is

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda}(\psi+\gamma'\Sigma^{-1}\gamma)^{(d/2)-\lambda}}{(2\pi)^{d/2}|\Sigma|^{1/2}K_\lambda(\sqrt{\chi\psi})}$$

we can see that if there is no skewness or gamma = 0, the distribution reduces to the symmetric GH special case in example 3 before. We can also find the charactesric function of the GH using the same techique as before to yield

$$\phi_X(t) = E(e^{it'X}) = e^{it'\mu}\hat{H}(\tfrac{1}{2}t'\Sigma t - it'\gamma)$$

thus, we can compile the notation of $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$. One note for the reader is that $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ are identical to $X \sim GH_d(\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ for $k > 0$, which eventually will cause an identifiability problem when we attempt to estimate the parameters in practice. We will see the solution to this problem in next section.

Parametrizations To solve the problem arises before, we need to renamed Σ to Δ and add a constraint such that $|\Delta| = 1$. The skewness parameters γ are replaced by parameters

β, and the non-negative parameters χ and ψ are replaced by the non-negative parameters δ and α according to

$$\beta = \Delta^{-1}\gamma,\ \delta = \sqrt{\chi},\ \alpha = \sqrt{\psi + \gamma'\Delta^{-1}\gamma}$$

where this parameters must satisfy the constraints $\delta \geq 0, \alpha^2 > \beta'\Delta\beta$ if $\lambda > 0$; $\delta > 0, \alpha^2 > \beta'\Delta\beta$ if $\lambda = 0$; and $\delta > 0, \alpha^2 \geq \beta'\Delta\beta$ if $\lambda < 0$.

3 Spherical and Elliptical Distributions

The normal variance mixture distributions perform better than the multivariate normal for the daily and weekly US stock-return data. However, the mean-variance mixture distributions may not be a good generalization for the normal variance mixture distributions. Then, we should consider the normal variance mixture distributions, which are often called as elliptical distributions, as our model. In the first place, we should start from a special case of spherical distributions.

3.1 Spherical Distributions

Definition 3.1.1. *A random vector $\boldsymbol{X} = (X_1, X_2, \cdots, X_d)'$ has a spherical distribution if, for every orthogonal map $U \in \mathbb{R}^{d\times d}$ (i.e. maps satisfying $UU' = U'U = I_d$),*

$$U\boldsymbol{X} \stackrel{d}{=} \boldsymbol{X}$$

By the definition, we know that spherical random vectors are distributional invariant under rotations. And we can define the distributions with this property by different ways. The following theorem states the details.

Theorem 3.1.1. *The followings are equivalent.*

(1) $\boldsymbol{X}$ is spherical.

(2) There exists a function ψ of a scalar variable such that, for all $\boldsymbol{t} \in \mathbb{R}^d$,

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}'\boldsymbol{X}}) = \psi(\boldsymbol{t}'\boldsymbol{t}) = \psi(t_1^2 + \cdots + t_d^2) \tag{3.1}$$

(3) For every $\boldsymbol{a} \in \mathbb{R}^d$,

$$\boldsymbol{a}'\boldsymbol{X} \overset{d}{=} ||\boldsymbol{a}||X_1 \tag{3.2}$$

where $||\boldsymbol{a}||^2 = \boldsymbol{a}'\boldsymbol{a} = a_1^2 + \cdots + a_d^2$

Proof. (1) $\Rightarrow$ (2). If $\boldsymbol{X}$ is spherical, then for any orthogonal matrix U we have

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \phi_{U\boldsymbol{X}}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}'U\boldsymbol{X}}) = \phi_{\boldsymbol{X}}(U'\boldsymbol{t}).$$

This can be true if $\phi_{\boldsymbol{X}}(\boldsymbol{t})$ only depends on the length of $\boldsymbol{t}$, i.e. if $\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \psi(\boldsymbol{t}'\boldsymbol{t})$ for some function ψ of a non-negative scalar variable.
(2) $\Rightarrow$ (3). First observe that $\phi_{X_1}(t) = E(e^{itX_1}) = \phi_{\boldsymbol{X}}(t\boldsymbol{e}_1) = \psi(t^2)$, where $\boldsymbol{e}_1$ denotes the first unit vector in $\mathbb{R}^d$. It follows that for any $\boldsymbol{a} \in \mathbb{R}^d$,

$$\phi_{\boldsymbol{a}'\boldsymbol{X}}(t) = \phi_{\boldsymbol{X}}(t\boldsymbol{a}) = \psi(t^2\boldsymbol{a}'\boldsymbol{a}) = \psi(t^2||\boldsymbol{a}||^2) = \phi_{X_1}(t||\boldsymbol{a}||) = \phi_{||\boldsymbol{a}||X_1}(t).$$

(3) $\Rightarrow$ (1). For any orthogonal matrix U we have

$$\phi_{U\boldsymbol{X}}(\boldsymbol{t}) = E(e^{i(U'\boldsymbol{t})'\boldsymbol{X}}) = E(e^{i||U'\boldsymbol{t}||X_1}) = E(e^{i||\boldsymbol{t}||X_1}) = E(e^{i\boldsymbol{t}'\boldsymbol{X}}) = \phi_{\boldsymbol{X}}(\boldsymbol{t}).$$

□

From the part (2) of Theorem 3.1.1 we know that the characteristic function of a spherically distributed random vector is fully described by a function ψ of a scalar variable. Then ψ is known as the characteristic generator of the spherical distribution and we use the notation $\boldsymbol{X} \sim S_d(\psi)$. Moreover, part (3) of the Theorem 3.1.1 shows that linear combinations of

spherical random vectors always have a distribution of the same type, so that they have the same distribution up to changes of location and scale. Next, an important way of characterizing spherical distribution is given by the following result.

Theorem 3.1.2. $\boldsymbol{X}$ *has a spherical distribution if and only if it has the stochastic representation*

$$\boldsymbol{X} \stackrel{d}{=} R\boldsymbol{S} \tag{3.3}$$

where $\boldsymbol{S}$ *is uniformly distributed on the unit sphere* $\mathfrak{s}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \boldsymbol{s}'\boldsymbol{s} = 1\}$ *and* $R \geq 0$ *is a radial r.v., independent of* $\boldsymbol{S}$.

Proof. ($\Leftarrow$) We consider the characteristic function

$$\phi_{R\boldsymbol{S}}(\boldsymbol{t}) = E(e^{iR\boldsymbol{t}'\boldsymbol{S}}) = E(E(e^{iR\boldsymbol{t}'\boldsymbol{S}}|R)).$$

Since $\boldsymbol{S}$ is spherically distributed, its characteristic function has a characteristic generator, which is usually given the special notation Ω_d. Thus, by the part (2) of Theorem 3.1.1 we have that

$$\phi_{R\boldsymbol{S}}(\boldsymbol{t}) = E(e^{iR\boldsymbol{t}'\boldsymbol{S}}) = E(E(e^{iR\boldsymbol{t}'\boldsymbol{S}}|R)) = E(\Omega_d(R^2\boldsymbol{t}'\boldsymbol{t})) = \int \Omega_d(r^2\boldsymbol{t}'\boldsymbol{t})dF(r), \tag{3.4}$$

where F is the df of R. Since this is a function of $\boldsymbol{t}'\boldsymbol{t}$, it follows from the part (2) of Theorem 3.1.1 that $R\boldsymbol{S}$ has a spherical distribution.

($\Rightarrow$) For $\forall \boldsymbol{s} \in \mathfrak{s}^{d-1}$, the characteristic generator ψ of spherical random variable $\boldsymbol{X}$ must satisfy $\psi(\boldsymbol{t}'\boldsymbol{t}) = \phi_{\boldsymbol{X}}(\boldsymbol{t}) = \phi_{\boldsymbol{X}}(||\boldsymbol{t}||\boldsymbol{s})$. It follows that, if we introduce a random vector $\boldsymbol{S}$ that is uniformly distributed on the sphere $\mathfrak{s}^{d-1}$, we have

$$\psi(\boldsymbol{t}'\boldsymbol{t}) = \int_{\mathfrak{s}^{d-1}} \phi_{\boldsymbol{X}}(||\boldsymbol{t}||\boldsymbol{s})dF_{\boldsymbol{S}}(\boldsymbol{s}) = \int_{\mathfrak{s}^{d-1}} E(e^{i||\boldsymbol{t}||\boldsymbol{s}'\boldsymbol{X}})dF_{\boldsymbol{S}}(\boldsymbol{s}).$$

Interchanging the order of integration and using the Ω_d notation for the characteristic generator of $\boldsymbol{S}$ we have

$$\psi(\boldsymbol{t}'\boldsymbol{t}) = E(\Omega_d(||\boldsymbol{t}||^2||\boldsymbol{X}||^2)) = \int \Omega_d(\boldsymbol{t}'\boldsymbol{t}r^2)dF_{||\boldsymbol{X}||}(r). \tag{3.5}$$

where $F_{||\boldsymbol{X}||}$ is the df of $||\boldsymbol{X}||$. By comparison with (3.4) we see that (3.5) is the characteristic function of $R\boldsymbol{S}$, where R is and r.v. with df $F_{||\boldsymbol{X}||}$ that is independent of $\boldsymbol{S}$. □

In most cases, we only consider spherical random variables $\boldsymbol{X}$ in the subclass $S_d^+(\psi)$ for which $P(\boldsymbol{X} = \boldsymbol{0}) = 0$. Next, we have a useful corollary of Theorem 3.1.1.

Corollary 3.1.1. *Suppose* $\boldsymbol{X} \stackrel{d}{=} R\boldsymbol{S} \sim S_d^+(\psi)$. *Then,*

$$(||\boldsymbol{X}||, \frac{\boldsymbol{X}}{||\boldsymbol{X}||}) \stackrel{d}{=} (R, \boldsymbol{S}) \tag{3.6}$$

Proof. Let $f_1(\boldsymbol{x}) = ||\boldsymbol{x}||$ and $f_2(\boldsymbol{x}) = \frac{\boldsymbol{x}}{||\boldsymbol{x}||}$. It follows from (3.3) that

$$(||\boldsymbol{X}||, \tfrac{\boldsymbol{X}}{||\boldsymbol{X}||}) = (f_1(\boldsymbol{X}), f_2(\boldsymbol{X})) \stackrel{d}{=} (f_1(R\boldsymbol{S}), f_2(R\boldsymbol{S})) = (R, \boldsymbol{S}).$$

□

We will provide an example using the fact from the corollary (3.1.1).

Example 3.1.1. *Suppose* $\boldsymbol{X} \sim N_d(\boldsymbol{0}, I_d)$, *then* $\boldsymbol{X}$ *is a spherical random variable. Since* $\boldsymbol{X}'\boldsymbol{X} \sim \chi_d^2$, *it follows from (3.6) that* $R^2 \sim \chi_d^2$. *Note that* $\boldsymbol{X} \stackrel{d}{=} R\boldsymbol{S}$ *does not guarantee* $\boldsymbol{X}'\boldsymbol{X} \stackrel{d}{=} \boldsymbol{S}'R'R\boldsymbol{S} = R^2$, *we can only use corollary (3.1.1) to show. Now, we can use previous result to calculate* $E(\boldsymbol{S})$ *and* $cov(\boldsymbol{S})$.

$$\boldsymbol{0} = E(\boldsymbol{X}) = E(R)E(\boldsymbol{S}) \Rightarrow E(\boldsymbol{S}) = \boldsymbol{0},$$

$$I_d = cov(\boldsymbol{X}) = E(R^2)cov(\boldsymbol{S}) \Rightarrow cov(\boldsymbol{S}) = \frac{I_d}{d}. \tag{3.7}$$

Since $E(R^2) = d$ *when* $R^2 \sim \chi_d^2$.

Example 3.1.2. *Suppose* $\boldsymbol{X}$ *has a spherical normal variance mixture distribution* $\boldsymbol{X} \sim M_d(\mathbf{0}, I_d, \hat{H})$. *We want to find the distribution of* $R^2 \stackrel{d}{=} \boldsymbol{X}'\boldsymbol{X}$ *in this case.*

By the definition,

$$\boldsymbol{X} \stackrel{d}{=} \sqrt{W}\boldsymbol{Y},$$

where $\boldsymbol{Y} \sim N_d(\mathbf{0}, I_d)$ *and* W *is independent of* $\boldsymbol{Y}$.

Next, since

$$R^2 \stackrel{d}{=} W\tilde{R}^2,$$

where $\tilde{R}^2 \stackrel{d}{=} \chi^2_d$ *and* W *is independent of* $\tilde{R}$. *If we can find the distribution of the product of* W *and an independent chi-squared random variable, then we can find the distribution of* R^2.

More specifically, if $\boldsymbol{X} \sim t_d(v, \mathbf{0}, I_d)$. *For a multivariate t distribution we know from previous example that* $W \sim Ig(\frac{1}{2}v, \frac{1}{2}v)$, *which means that* $v/W \sim \chi^2_v$.

Since we know the fact that ratio of independent chi-squared random variables divided by their degrees of freedom is F*-distribution, we have*

$$\frac{\frac{\chi^2_d}{d}}{\frac{\chi^2_v}{v}} \sim F(d, v)$$

$$\frac{\frac{\chi^2_d}{d}}{\frac{\chi^2_v}{v}} = \frac{\frac{\tilde{R}^2}{d}}{\frac{\frac{W}{v}}{v}} = \frac{W\tilde{R}^2}{d} = \frac{R^2}{d}$$

$$\Rightarrow \frac{R^2}{d} \sim F(d, v)$$

which is a F *distribution on* d *with* v *degrees of freedom. Since the mean of* $F(d, v)$ *distribution is* $\frac{v}{v-2}$, *it follows from (3.7) that*

$$cov(\boldsymbol{X}) = E(cov(R\boldsymbol{S}|R)) = E(R^2\tfrac{I_d}{d}) = (\tfrac{v}{v-2})I_d.$$

3.2 Elliptical Distributions

Definition 3.2.1. $\boldsymbol{X}$ *has an elliptical distribution if*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Y}$$

where $\boldsymbol{Y} \sim S_k(\psi)$ *and* $A \in \mathbb{R}^{d\times k}$ *and* $\mu \in \mathbb{R}^d$ *are a matrix and vector of constants.*

Then the elliptical distributions are obtained by multivariate affine transformations of spherical distributions. So we have,

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = E(e^{i\boldsymbol{t}'\boldsymbol{X}}) = E(e^{i\boldsymbol{t}'(\boldsymbol{\mu}+A\boldsymbol{Y})}) = e^{i\boldsymbol{t}'\boldsymbol{\mu}}E(e^{i(A'\boldsymbol{t})'\boldsymbol{Y}}) = e^{i\boldsymbol{t}'\boldsymbol{\mu}}\psi(\boldsymbol{t}'\Sigma\boldsymbol{t})$$

where $\Sigma = AA'$, we denote the elliptical distribution by

$$\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$$

and refer to $\boldsymbol{\mu}$ as the location vector, Σ as the dispersion matrix and ψ as the characteristic generator of the distribution.

Note that $\boldsymbol{X}$ does not uniquely determine its elliptical representation by $E_d(\boldsymbol{\mu}, \Sigma, \psi)$. Although $\boldsymbol{\mu}$ is uniquely determined, Σ and ψ are only determined up to a positive constant. For example, the multivariate normal distribution $N_d(\boldsymbol{\mu}, \Sigma)$ can be written as $E_d(\boldsymbol{\mu}, \Sigma, \psi(\cdot))$ or $E_d(\boldsymbol{\mu}, \Sigma, \psi(\frac{\cdot}{c}))$ for $\psi(u) = e^{-\frac{u}{2}}$ and any $c > 0$. Provided that variances are finite, then an elliptical distribution is fully specified by its mean vector, covariance matrix and characteristic generator, and it is possible to find an elliptical representation $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ such that Σ is the covariance matrix of $\boldsymbol{X}$, although this is not always the standard representation of the distribution.

With the same manner, we give an alternative stochastic representation for the elliptical distributions that follows directly from Definition 3.2.1 and Theorem 3.1.2.

Proposition 3.2.1. $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ *if and only if there exists* $\boldsymbol{S}$*,* R *and* A *satisfying*

$$\boldsymbol{X} \stackrel{d}{=} \boldsymbol{\mu} + RA\boldsymbol{S} \tag{3.8}$$

with

1. $\boldsymbol{S}$ *uniformly distributed on the unit sphere* $\mathfrak{s}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \boldsymbol{s}'\boldsymbol{s} = 1\}$,

2. $R \geq 0$, *a radial random variable, independent of* $\boldsymbol{S}$, *and*

3. $A \in \mathbb{R}^{d\times k}$ *with* $AA' = \Sigma$.

For practical examples we are most interested in the case where Σ is positive definite. The relation between the elliptical and spherical cases is then clearly

$$\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi) \Leftrightarrow \boldsymbol{Y} = \Sigma^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu}) \sim S_d(\psi). \tag{3.9}$$

In this case, if the spherical vector $\boldsymbol{Y}$ has density generator g, then $\boldsymbol{X} = \boldsymbol{\mu} + \Sigma^{\frac{1}{2}}\boldsymbol{Y}$ has density

$$f(\boldsymbol{X}) = \tfrac{1}{|\Sigma|^{\frac{1}{2}}} g((\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})).$$

The joint density is always constant on sets of the form $\{\boldsymbol{x} : (\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = c\}$, which are ellipsoids in $\mathbb{R}^d$. Clearly, the full family of multivariate normal variance mixtures with general location and dispersion parameters $\boldsymbol{\mu}$ and Σ are elliptical, since they are obtained by affine transformations of the spherical special cases considered in the previous section.

3.3 Properties of Elliptical Distributions

Now, we summarize some of the properties of elliptical distributions. It turns out that we can carry over most of the properties from multivariate normal distributions directly. These parallels emphasize that it would be easy to base many standard procedures in risk management on an assumption that risk-factor changes have an approximately elliptical distribution, rather than the false assumption that they are multivariate normal.

1. Linear Combination Let $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and take any $B \in \mathbb{R}^{k\times d}$ and $\boldsymbol{b} \in \mathbb{R}^k$. We

know that

$$B\boldsymbol{X} + \boldsymbol{b} \sim E_k(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B^{'}, \psi). \tag{3.10}$$

As a special case, if $\boldsymbol{a} \in \mathbb{R}^d$, then

$$\boldsymbol{a}^{'}\boldsymbol{X} \sim E_1(\boldsymbol{a}^{'}\boldsymbol{\mu}, \boldsymbol{a}^{'}\Sigma\boldsymbol{a}, \psi). \tag{3.11}$$

2. Marginal Distributions It follows from (3.11) that marginal distribution of $\boldsymbol{X}$ must be elliptical distributions with the same characteristic generator. Using the same $\boldsymbol{X} = (\boldsymbol{X}_1^{'}, \boldsymbol{X}_2^{'})$ notation as previous section and again extending this notation naturally to $\boldsymbol{\mu}$ and Σ, we have that $\boldsymbol{X}_1 \sim E_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and $\boldsymbol{X}_2 \sim E_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$.

3. Conditional Distributions The conditional distribution of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1$ may also be shown to be elliptical, although in general it will have a different characteristic generator $\tilde{\psi}$. For details of how the generator changes see Fang, Kotz and Ng (1990, pp. 45, 46). In the special case of multivariate normality the generator remains the same.

4. Quadratic Form If $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with Σ non-singular, then

$$Q := (\boldsymbol{X} - \boldsymbol{\mu})^{'}\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \stackrel{d}{=} R^2 \tag{3.12}$$

where R is the radial random variable in the stochastic representation (3.8). As we have seen in Example 3.1.1, for some particular cases the distribution of R^2 is well known: if $\boldsymbol{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, then $R^2 \sim \chi_d^2$; if $\boldsymbol{X} \sim t_d(v, \boldsymbol{\mu}, \Sigma)$, then $R^2 \sim F(d, v)$. For all elliptical distributions, Q must be independent of $\frac{\Sigma^{-\frac{1}{2}}(\boldsymbol{X}-\boldsymbol{\mu})}{\sqrt{Q}}$.

5. Convolutions The convolution of two independent elliptical vectors with the same dispersion matrix Σ is also elliptical. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent d-dimensional random vectors satisfying $\boldsymbol{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\boldsymbol{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, \Sigma, \tilde{\psi})$, then we may take the product of characteristic functions to show that

$$\boldsymbol{X} + \boldsymbol{Y} \sim E_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma, \varphi) \tag{3.13}$$

where $\varphi(u) = \psi(u)\tilde{\psi}(u)$. Moreover, if the dispersion matrices of $\boldsymbol{X}$ and $\boldsymbol{Y}$ differ by more than a constant factor, then the convolution will not necessarily remain elliptical, even when the two generators ψ and $\tilde{\psi}$ are identical.

3.4 Estimating Dispersion and Correlation

For now, we know that risk-factor return data $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ are from some elliptical distribution $E_d(\boldsymbol{\mu}, \Sigma, \psi)$ with heavier tails than the multivariate normal. In this section, we want to estimate the location parameter $\boldsymbol{\mu}$, a dispersion matrix Σ and the correlation matrix P, assuming finite second moments. We could use the standard estimators. Even though they are unbiased and consistent under quite weak assumptions, they may not be the best estimators of location and dispersion. There are many alternative estimators that may be more efficient for heavy-tailed data and may have better robustness properties for contaminated data. Firstly, we introduce Maronna's M-estimators for estimating location and dispersion parameters.

Let $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ denote estimates of the mean vector and the dispersion matrix. Suppose for every observation $\boldsymbol{X}_i$ we calculate $D_i^2 = (\boldsymbol{X}_i - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}_i - \boldsymbol{\mu})$. If we wanted to improve estimates of location and dispersion, particularly for heavy-tailed data,we need to reduce the influence of observations for large D_i, since these are the observations that might tend to distort the estimation of the parameters. Thus, we will define a decreasing weight function $w_j : \mathbb{R}^+ \to \mathbb{R}^+$, $j = 1, 2$, to reduce the influence of observations with large D_i. This can be turned into an iterative procedure that converges to so-called M-estimates of location and dispersion; the dispersion matrix estimate is generally a biased estimate of the true covariance matrix.

Algorithm 3.4.1. ***(M-estimators of location and dispersion)***.

(1) As starting estimates take $\hat{\boldsymbol{\mu}}^{[1]} = \bar{X}$ *and* $\hat{\Sigma}^{[1]} = S$*, the standard estimators in (6.9). Set*

iteration count $k = 1$.

(2) For $i = 1, 2, \cdots, n$, *set* $D_i^2 = (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})' \hat{\Sigma}^{[k]-1} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})$.

(3) Update the location estimate using

$$\hat{\boldsymbol{\mu}}^{[k+1]} = \frac{\sum_{i=1}^{n} w_1(D_i) \boldsymbol{X}_i}{\sum_{i=1}^{n} w_1(D_i)}.$$

where w_1 *is a weight function, as discussed below.*

(4) Update the dispersion matrix estimate using

$$\hat{\Sigma}^{[k]+1} = \frac{1}{n} \sum_{i=1}^{n} w_2(D_i^2 (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})(\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}^{[k]})')$$

where w_2 *is a weight function.*

(5) Set $k = k + 1$ *and repeat steps (2)-(4) until estimates converge.*

There are many weight functions. One would be the decreasing functions $w_1(x) = w_2(x^2) = \frac{d+v}{x^2+v}$ for some positive constant v. Use of these weight functions in Algorithm 3.4.1 exactly corresponds to fitting a multivariate $t_d(v, \boldsymbol{\mu}, \Sigma)$ distribution with known degrees of freedom v using the EM algorithm. For more details, see Meng and van Dyk (1997). The other choice would be a function such that

$$w_1(x) = 1, \; x \leq a$$
$$w_1(x) = \frac{a}{x}, \; x > a$$

for some a and

$$w_2(x^2) = (w_1(x))^2$$

In this case, we consider a function that gives full weight for the central part of the distribution and the decreasing weight for the outliers.

Secondly, we introduce a method to estimate the correlation based on Kendall's rank correlation coefficient. The theoretical version of Kendall's rank correlation (also known

as Kendall's tau) for two random variables X_1 and X_2 is denoted by $\rho_\tau(X_1, X_2)$. And if $(X_1, X_2) \sim E_2(\boldsymbol{\mu}, \Sigma, \psi)$, then

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi} arcsin(\rho) \tag{3.14}$$

where $\rho = \sigma_{12}/(\sigma_{11}\sigma_{22})^{1/2}$ is the pseudo-correlation coefficient of the elliptical distribution, which is always defined (even when correlation coefficients are undefined because variances are infinite). By inverting this relationship, we can get a method for estimating ρ from data. In the first place, we can find the estimator of Kendall's tau $\hat{\rho}_\tau(X_1, X_2)$. Then, we have an estimator for $\hat{\rho}$

Note that this method can be used to estimate a correlation matrix of a higher-dimensional elliptical distribution by applying the technique to each bivariate margin. This may result in a matrix of pairwise correlation estimates that is not necessarily positive definite; this problem does not always arise, and if it does, a matrix adjustment method can be used, more details can be found from the eigenvalue method of Rousseeuw and Molenberghs (1993).

Moreover, if we want to turn an estimate of a bivariate correlation matrix into a robust estimate of a dispersion matrix we could estimate the ratio of standard deviations $\lambda = (\sigma_{11}\sigma_{22})^{1/2}$ by using a ratio of trimmed sample standard deviations. In other words, we drop an equal number of outliers from each of the univariate data sets $X_{1,i}, \cdots, X_{n,i}$ for $i = 1, 2$ and calculate the sample standard deviations with the remaining observations. We could have the estimate as

$$\hat{\Sigma} = \begin{pmatrix} 1 & \hat{\lambda}\hat{\rho} \\ \hat{\lambda}\hat{\rho} & \hat{\lambda}^2 \end{pmatrix} \tag{3.15}$$

4 Dimension-Reduction Techniques

As we have been discussing a lot about multivariate distributions, we should realize that the complexity of a real-world problem can grow rapidly as the number of dimensions increases. For example, a computer algorithm designed for multivariate statistical analysis may require more running time and space for a higher dimensional problem. To address this issue, dimension-reduction techniques, such as factor models and principal component analysis (PCA), are powerful tools and widely used in practice to simplify the problem.

4.1 Linear Factor Models

Definition 4.1.1. *In linear factor models, the randomness in multivariate data is explained by a set of common factors with fewer dimensions. A factor model also includes a vector of constants and a vector of error terms, thus taking the form of*

$$X = a + BF + \varepsilon, \tag{4.1}$$

where

(i) $X = (X_1, X_2, ..., X_d)'$ *is a d-dimensional vector, of which the number of dimensions needs to be reduced;*

(ii) $a \in \mathbb{R}^d$ *is a vector of constants;*

(iii) $F = (F_1, F_2, ..., F_p)'$ *is a p-dimensional* $(p < d)$ *multivariate random variable with a positive definite covariance matrix;*

(iv) $B \in \mathbb{R}^{d \times p}$ *is a matrix of constant factor loadings; and*

(v) $\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_d)'$ *is a d-dimensional random vector of idiosyncratic error terms with a mean of 0.*

Assumptions. Two important underlying assumptions of factors models are that

(i) The multivariate random vector of common factors F is uncorrelated with the error

terms ε, i.e. $cov(F, \varepsilon) = 0$; and

(ii) The errors are uncorrelated with each other.

As mentioned above, the purpose of factor modeling is to explain the randomness in X, represented by its covariance matrix Σ, thus, we are able to decompose Σ into the variations of the two random vectors on the right-hand side of equation 4.1, and we have the following relationship

$$\Sigma = B\Omega B' + \Upsilon, \tag{4.2}$$

where Ω is the covariance matrix of F, and Υ is the covariance matrix of ε. Υ is apparently diagonal because the pairwise correlation of every two different components of ε is 0. Because the covariance matrix of F is positive definite, we can set $B^* = B\Omega^{1/2}$ and rewrite equation 4.2 as

$$\Sigma = B^*(B^*)' + \Upsilon. \tag{4.3}$$

Equicorrelation model

A special case of this type of factor models, which is widely used in practice, is the equicorrelation model, wherein the pairwise correlations of the components in X are all equal with a value of ρ. For simplicity, we let each component of X be standard normal, thus, the covariance matrix Σ can be written as

$$\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix},$$

or equivalently, $\Sigma = \rho J_d + (1-\rho) I_d$, where J_d is a d-dimensional square matrix with all elements equal to 1, and I_d is the d-dimensional identity matrix. It is easy to recognize that this equation takes the same form as equation 4.3. In this case, $B^*(B^*)' = \rho J_d$, that is to say, $B^* = \sqrt{\rho}1$, which is a d-dimensional vector with elements equal to $\sqrt{\rho}$. Also, Υ,

the diagonal covariance matrix of ε, is equal to $(1-\rho)I_d$, suggesting that the variance of a single error term is $(1-\rho)$.

One way of decomposing X is to set the common factor F, which in this case is one-dimensional, equal to

$$F = \frac{\sqrt{\rho}}{1+\rho(d-1)} \sum_{j=1}^{d} X_j + \sqrt{\frac{1-\rho}{1+\rho(d-1)}} Y, \tag{4.4}$$

where Y can be any random variable independent of X with zero mean and unit variance. In this setting, the vector of errors can be obtained by

$$\varepsilon = X - B^* F, \tag{4.5}$$

and because $B^* = \sqrt{\rho}1$, individual elements in this vector can be represented by

$$\varepsilon_j = X_j - \sqrt{\rho} F. \tag{4.6}$$

We can verify that the F here is a random variable with a mean of 0 and a variance of 1.

$$\begin{aligned} E[F] &= \frac{\sqrt{\rho}}{1+\rho(d-1)} E[\sum_{j=1}^{d} X_j] + \sqrt{\frac{1-\rho}{1+\rho(d-1)}} E[Y] \\ &= 0, \end{aligned}$$

$$\begin{aligned} Var(F) &= \frac{\rho}{[1+\rho(d-1)]^2} Var[\sum_{j=1}^{d} X_j] + \frac{1-\rho}{1+\rho(d-1)} Var[Y] \\ &= \frac{\rho(d+d(d-1)\rho)}{[1+\rho(d-1)]^2} + \frac{1-\rho}{1+\rho(d-1)} \\ &= \frac{1+\rho(d-1)}{1+\rho(d-1)} \\ &= 1 \end{aligned}$$

We can further verify that F and ε_j are uncorrelated and that errors are uncorrelated

with each other, which means that this is a valid factor model.

$$\begin{aligned} cov(F, \varepsilon_j) &= E[F\varepsilon_j] - E[F]E[\varepsilon_j] \\ &= E[F\varepsilon_j] \\ &= E[F(X_j - \sqrt{\rho}F)] \\ &= E[FX_j] - \sqrt{\rho} \\ &= \frac{\sqrt{\rho}}{1+\rho(d-1)} E[X_j \sum_{i=1}^{d} X_i] - \sqrt{\rho} \\ &= \frac{\sqrt{\rho}(1+\rho(d-1))}{1+\rho(d-1)} - \sqrt{\rho} = 0, \end{aligned}$$

$$\begin{aligned} cov(\varepsilon_i, \varepsilon_j, i \neq j) &= E[\varepsilon_i\varepsilon_j] - E[\varepsilon_i]E[\varepsilon_j] \\ &= E[\varepsilon_i\varepsilon_j] \\ &= E[\varepsilon_i(X_j - \sqrt{\rho}F)] \\ &= E[\varepsilon_i X_j] \\ &= E[(X_i - \sqrt{\rho}F)X_j] \\ &= \rho - (\sqrt{\rho})^2 = 0 \end{aligned}$$

Because the random variable Y in F can follow any zero-mean, unit-variance distribution, we can let it be normally distributed, which makes both F and ε_j Gaussian under the condition that X is also Gaussian. As we have shown that the variance of ε_j is $(1-\rho)$, and the constant factor loading of F is $\sqrt{\rho}$, we can rewrite the factor model in the form of

$$X_j = \sqrt{\rho}F + \sqrt{1-\rho}Z_j, \quad j = 1, 2, ..., d, \tag{4.7}$$

where F and Z_j are independent standard normal random variables. This equation essentially represents a 1-factor model and provides a convenient way to sample from a multivariate normal distribution with desired correlation and number of dimensions.

One-factor Gaussian copula model

One application of this model is estimating the credit risk of bond portfolios. We can conceptualize that in a bond portfolio, all obligors are subject to the systematic risk from the market, whereas each of them also has its own idiosyncratic risk. Different obligors may default at the same time when the market performs poorly, that is to say, their default times (or times to default) may be correlated. This scenario can be nicely represented by equation 4.7, wherein the common factor F can be seen as a risk factor driven by the market, and Z_j represents the idiosyncratic risk of each obligor.

For illustration, suppose in a bond portfolio, there are n obligors, and their default times T_i follow the exponential distribution $F_i(t) = P(T_i \leq t) = 1 - e^{\lambda_i t}, i = 1, 2, ..., n$, where λ_i is the probability of default of each obligor. We further know that the correlation of their default times is estimated to be ρ. In order to simulate one observation from their joint distribution with the desired correlation structure, we can use the following algorithm based on a 1-factor Gaussian copula model:

Algorithm 4.1.1. *Modeling correlated default times with 1-factor Gaussian copula*

i. Simulate a market component m from N(0,1).

ii. Simulate n idiosyncratic risk components, $\varepsilon_i, i = 1, 2, .., n$, from N(0,1) separately.

iii. By setting $x_i = \sqrt{\rho}m + \sqrt{1-\rho}\varepsilon_i$, we get a sample of size 1 from a standard multivariate normal distribution with correlation ρ.

iv. Transform x_i to a uniform random variable u_i using $u_i = \Phi(x_i)$.

v. u_i represents the percentile of default time T_i, and we can get the corresponding quantile t_i, which is the simulated default time of obligor i, from $F_i^{-1}(u_i) = -ln(1 - u_i)/\lambda_i$.

The pair of default times (t_1, t_2) is a sample of size 1 drawn from the multivariate joint

distribution with exponential margins and correlation ρ. A numerical example of this method is given as follows.

Example 4.1.1. *Two obligors in a bond portfolio have probabilities of default* $\lambda_1 = 0.05, \lambda_2 = 0.1$*, and their default times are correlated by* $\rho = 0.7$*. Using this configuration, we produce a sample of default times of size 1000 using the above-mentioned algorithm. The left panel shows a bivariate normal sample of size 1000 created using the one-factor model, and the right panel shows the correlated default times of these two obligors produced by Gaussian copula. Because obligor 1 has a lower probability of default, thus, its default times are generally longer than the default times of obligor 2. Meanwhile, their correlation remains at 0.7.*

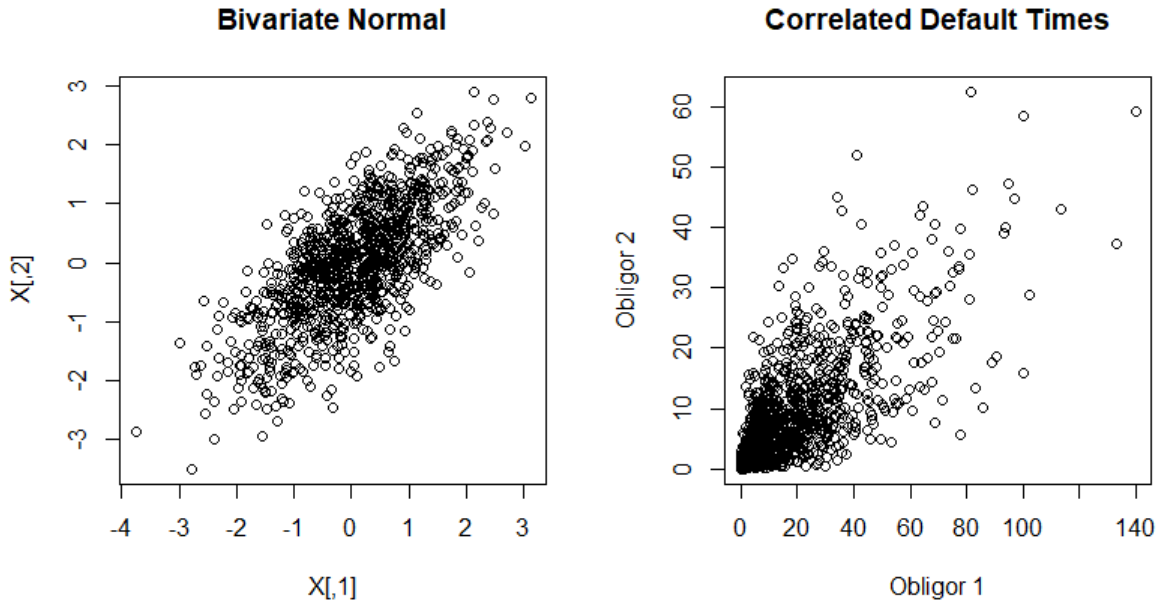

Figure 4.1: Left: standard bivariate normal data, sample $\rho = 0.7172$; Right: default times, sample $\rho = 0.7019$

Factor models can also be used to estimate parameters in real data. This approach can be seen as a generalization of univariate regression, and the difference is that the dependent

variable is multivariate. Commonly used strategies include macroeconomic factor models and fundamental factor models. In macroeconomic factor models, the common factors are numeric variables, which are proxies for the macroeconomic environment, such as stock indices and interest rates. In fundamental factor models, the common factors are categorical, indicating the underlying fundamentals of the studied subjects, such as country and business sector. In the Implementation and Application section of this article, an example of macroeconomic factor model applied on stock return data will be provided.

4.2 Principal Component Analysis

As these two strategies both require a set of handpicked factors, many other factors are dropped and not able to provide any explanatory power. To overcome this problem, principal component analysis (PCA) constructs a new set of common factors, called principal components, based on the existing factors by rotating the data. Although these PCs have the same dimensionality as the existing factors, they are ordered by explanatory power, so that we can choose to drop the least explanatory ones. Consider the standard bivariate normal example shown in figure 4.2.

The two normal random variables are configured to have a high correlation $\rho = 0.9$, and they are both shifted by 2, as shown in the left panel of the figure. After being rotated and centered to $(0, 0)$ by PCA, which is essentially a linear combination of these two random variables, as we can see from the right panel of the figure, the data is distributed along the horizontal axis, demonstrating that PC1 explains most of the variances in the original data, whereas there is little variation along the axis of PC2. In this case, if we wish to reduce the dimensionality of the original data, we can drop PC2, which relatively has little explanatory power.

The rotation of data can be achieved based on the Spectral Decomposition Theorem.

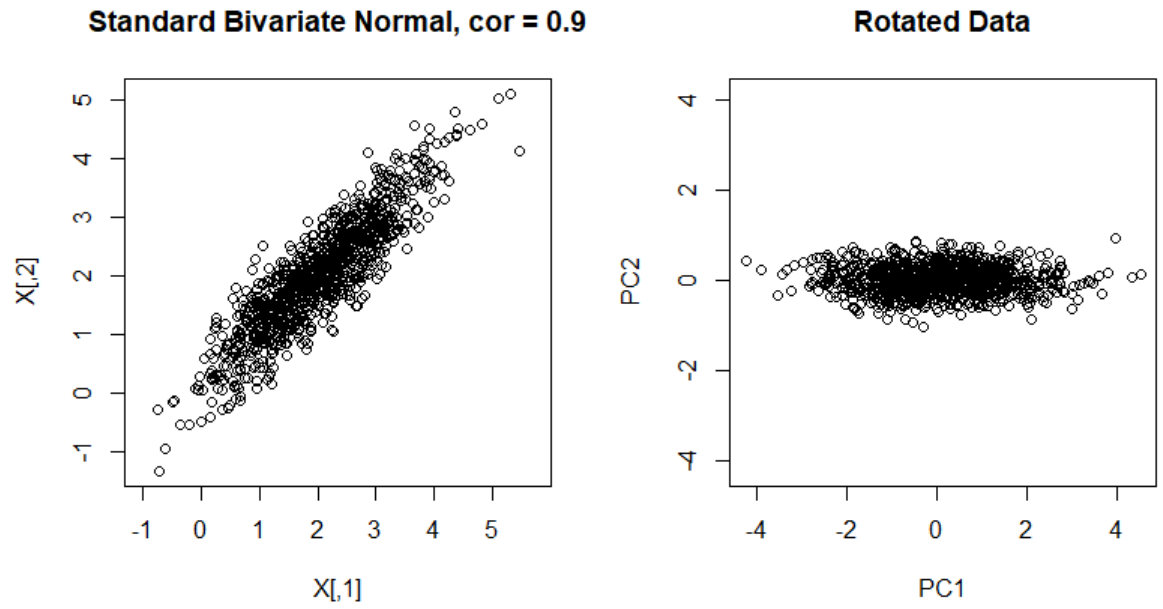

Figure 4.2: Left: standard bivariate normal data with $\rho = 0.9$; Right: rotated data using PCA

Theorem 4.2.1. *Spectral Decomposition Theorem*

Any symmetric matrix $A \in \mathbb{R}^{d \times d}$ *can be written as*

$$A = \Gamma \Lambda \Gamma', \tag{4.8}$$

where

i. $\Lambda = diag(\lambda_i, \lambda_2, ..., \lambda_n)$*, and* $\lambda_i, i = 1, 2, ..., d$ *are eigenvalues of matrix* A*, and*

ii. Γ *is a matrix wherein columns are eigenvectors corresponding to* λ_i*, and they constitute a orthonormal basis of* $\mathbb{R}^d$*, so that* $\Gamma\Gamma' = \Gamma'\Gamma = I_d$*.*

The matrix we need to decompose here is Σ, the covariance matrix of X. Since it is symmetric, it can be decomposed in this way. In addition, because we want to sort the PCs based on their explanatory power, the eigenvalues λ_i in the diagonal matrix Λ should

be put in a descending order, and the eigenvectors in Γ should be sorted in the same order as their corresponding eigenvalues. The matrix of eigenvalues Γ is the rotation matrix that performs the data transformation as we see in figure 4.2. Because PCA also involves a recentering process, aside from the rotation matrix Γ, a mean vector μ of X is also needed, the transform data Y can be obtained by

$$Y = \Gamma'(X - \mu) \tag{4.9}$$

which is the exact process that creates the right panel of figure 4.2, and each component in Y is a principal component. We can show that the diagonal matrix Λ is the covariance matrix of Y

$$cov(Y) = \Gamma' cov(X)\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda,$$

thus, principal components are uncorrelated, and each λ_i in Λ represents the variance of the corresponding principal component. In order to construct the factor model, we rewrite equation 4.9, so that X is a function of Y, and we get

$$X = \Gamma Y + \mu \tag{4.10}$$

$$X = \sum_{j=1}^{d} \Gamma_j Y_j + \mu. \tag{4.11}$$

In practice, we can reduce the dimensionality of X to a number $k < d$, by keeping the first k components as common factors and the remaining $d - k$ components as errors, so we have

$$X = \sum_{j=1}^{k} \Gamma_j Y_j + \varepsilon + \mu, \quad \varepsilon = \sum_{j=k}^{d} \Gamma_j Y_j \tag{4.12}$$

The variance of X explained by the common factors can be easily computed from $\sum_{j=1}^{k} \lambda_i$, since each λ_i is the variance of an associated principal component.

Part II

Implementation and Application

1 Example with Real Data

We took 7 years of data spanning the period 2010-2016 and formed daily, weekly, monthly and quarterly logarithmic returns. And it includes 10 stocks from AMZN, CPB, NVDA, CSCO, UNP, F, AIG, IBM, WMT and JCP, which are mostly large-cap companies, and they operate in diverse business sectors.

1.1 Test Normality of Returns on Stocks

In the first step, we apply Mardia's tests of multinormality based on both multivariate skewness and kurtosis to the multivariate data for all ten stocks. The results are shown in Table 5.1. We also compare the D_i^2 i data (1.18) to a χ^2_{10} 10 distribution using a QQ plot (see Figure 5.1). According to Table 5.1, for quarterly return data the multivariate kurtosis test and the skewness test does not reject the null hypothesis; the Q-Q plots for monthly returns and quarterly returns in Figure 5.1 look slightly more linear. There is therefore some evidence that returns over a month year and a quarter year are close to being normally distributed, which might indicate a central limit theorem effect taking place.

The results above are fairly typical for financial return data. This suggests that in many risk-management applications the multivariate normal distribution is not a good description of reality. It has three main defects:

1. The tails of its univariate marginal distributions are too thin; they do not assign enough weight to extreme events.

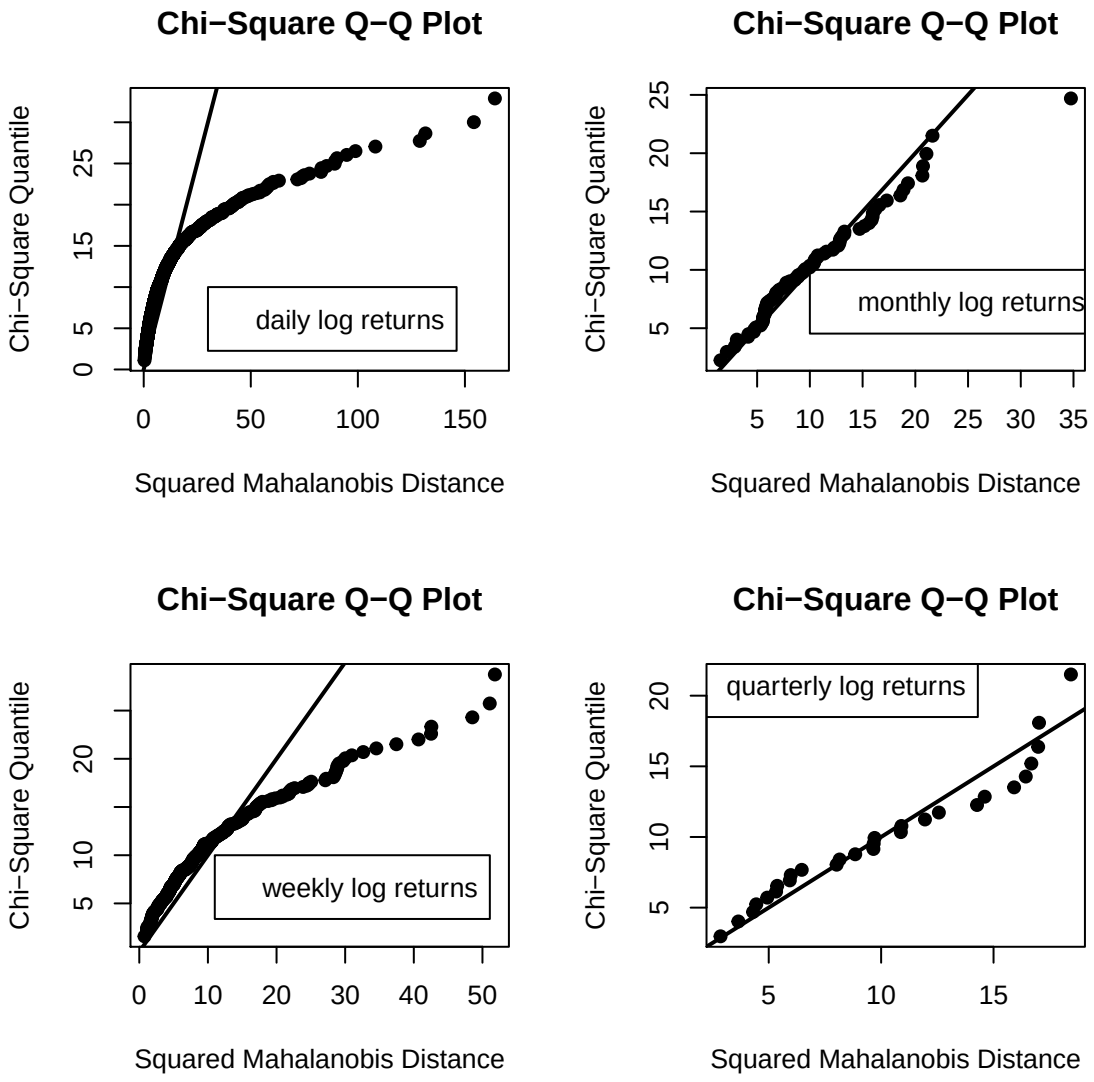

Figure 1.1: Under the null hypothesis of multivariatenormality these should be roughly linear.

2. The joint tails of the distribution do not assign enough weight to joint extreme outcomes.

3. The distribution has a strong form of symmetry, known as elliptical symmetry.

In the next step, we look at models that address some of these defects. We consider normal variance mixture models, which share the elliptical symmetry of the multivariate normal

Table 1.1: Tests of multinormality

Type	n	Test	Statistic	p value	Result
daily	1762	Mardia Skewness	3738.66125709258	0.000	NO
daily	1762	Mardia Kurtosis	209.591904344901	0.000	NO
weekly	365	Mardia Skewness	717.922619305109	0.000	NO
weekly	365	Mardia Kurtosis	29.2612257321467	0.000	NO
monthly	84	Mardia Skewness	264.785967195648	0.021	NO
monthly	84	Mardia Kurtosis	3.19527308412814	0.001	NO
quarterly	28	Mardia Skewness	252.258742926844	0.060	YES
quarterly	28	Mardia Kurtosis	0.362858710144666	0.717	YES

but have the flexibility to address (1) and (2) above; we also look at normal mean-variance mixture models, which introduce some asymmetry and thus address (3).

1.2 multivariate GH distribution

Now, we apply the multivariate GH distribution to the data and find a most suitable distribution for our data. We only use daily and weekly data because we know that they are not satisfied the normality assumption. Moreover, we exam whether we could replace the general mean-variance mixture by the variance mixtures (elliptical distribution).

univariate stock return

In the section, we fit our data by symmetric and asymmetric cases of the t, NIG, and hyperbolic models. We assume that the models are fitted using maximum likelihood under the simplifying assumption that returns form iid samples.

The table 5.2 shows the comparison of univariate models in the GH family, and it also shows the estimates of the selecting parameters and the value of the log-likelihood. Moreover, number in bold means the most suitable model which gives the largest value of the log-likelihood. The fist part of the table 5.2 discuss about the daily and symmetric case. For daily data, there are 7 out of 10 stocks returns prefer the t distribution among all the

Table 1.2: Comparison of univariate models in the GH family

	Gauss	t model		NIG model		Hyperbolic model	
Stock	ln L	v	ln L	$\sqrt{\chi\psi}$	ln L	$\sqrt{\chi\psi}$	ln L
Daily returns: symmetric models							
AIG	4257.5	2.6	4521.1	0.4	**4522.9**	0.2	4503.1
AMZN	4365.1	3.7	**4546.9**	0.8	4540.0	0.6	4528.5
CPB	5499.7	4.3	**5613.2**	1.0	5610.1	0.8	5605.1
CSCO	4726.3	3.1	**5066.2**	0.5	5054.0	0.2	5031.0
F	4550.3	3.9	4654.0	0.8	**4656.7**	0.5	4656.0
IBM	5275.9	3.9	**5412.4**	0.9	5409.6	0.6	5403.0
JCP	3572.9	3.8	**3737.7**	0.8	3732.2	0.6	3722.4
NVDA	4050.5	3.3	**4250.0**	0.6	4248.4	0.2	4241.0
UNP	4917.0	4.7	4977.7	1.2	**4980.2**	0.9	4980.1
WMT	5557.2	3.8	**5729.6**	0.8	5724.4	0.6	5715.6
Weekly returns: symmetric models							
AIG	607.6	2.8	638.8	0.5	**641.0**	0.2	640.2
AMZN	633.6	6.7	**641.2**	2.1	641.1	2.0	640.9
CPB	849.1	6.8	**856.7**	2.1	856.6	2.0	856.4
CSCO	695.6	3.9	**719.9**	0.8	719.9	0.5	719.4
F	644.6	6.2	650.1	1.7	650.9	1.2	**651.2**
IBM	809.5	4.5	823.5	1.1	**824.1**	0.8	823.9
JCP	416.9	3.8	438.7	0.8	**439.3**	0.5	438.7
NVDA	545.2	6.4	**555.6**	1.9	555.2	1.7	554.9
UNP	739.4	9.6	741.7	3.1	742.0	2.8	**742.1**
WMT	868.0	5.9	**881.4**	1.8	880.6	1.7	880.0
Weekly returns: asymmetric models							
WMT	NA	6.8	**883.7**	2.1	883.0	2.1	882.5

other members; for weekly data, the t distribution would again be the most suitable model. Overall, it turns out that t distribution would be a wise choice when we are fitting the stock returns for daily data. Moreover, all the mixture models fit much better than the Gaussian model in all cases.

For the asymmetric models, we only show cases where at least one of the asymmetric mixture models shows a significant improvement ($p < 0.05$) on the corresponding symmetric model according to a likelihood ratio test. It turns out that improvement only occurs for

weekly returns on Walmart (WMT). For Walmart the p-values of the tests were all 0.03 (t distribution, NIG, Hyperbolic).

multivariate stock return

Table 1.3: A comparison of models in the GH family for ten-dimensional stock-return data

	GH	NIG	Hyperbolic	t	VG	Gauss
			Daily returns: asymmetric models			
ln L	**50873.81**	50836.01	50453.27	**50873.81**	50726.04	
# par.	77	76	76	76	76	
p-value		0.00	0.00	1.00	0.00	
			Daily returns: symmetric models			
ln L	50870.32	50832.77	50447.56	50870.32	50723.07	48976.51
# par.	67	66	66	66	66	65
p-value	0.01	0.00	0.00	0.03	0.00	0.00
			Weekly returns: asymmetric returns			
ln L	**7422.47**	**7422.00**	7404.75	7421.82	7414.30	
p-value		0.34	0.00	0.26	0.01	
			Weekly returns: symmetric models			
ln L	7416.48	7415.91	7397.85	7415.95	7410.63	7265.93
p-value	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.3 shows the comparison of models in the GH family for ten-dimensional stock-return data. The p-values are for the likelihood ratio test of all other cases against the asymmetric GH model. For the daily data, the most suitable model is the skewed t distribution, which has a largest value for the maximum likelihood value and cannot be improved by the Gh. All other cases are rejected in a likelihood ratio test; for weekly data, the asymmetric NIG is the best model, followed closely by the asymmetric t distribution, and all other cases are rejected by the likelihood ratio test.

1.3 Dimension-Reduction Techniques

As we previously discussed about the macroeconomic factor model, in this section, we demonstrate this parameter estimation strategy using real stock return data, which includes

the daily logarithmic returns of 10 stocks from 2010 to 2016, totaling 1762 observations, denoted by X.

Macroeconomic Factor Model

For the macroeconomic factor, we use Russell 3000 index, which represents a mixture of large-cap and small-cap companies, and the model is construct as follows

$$X = BM + E,$$

where

i. X is a 1762×10 matrix with each column to be the returns of a single stock,

ii. B is the vector of constant factor loadings,

iii. M is a 1762×2 matrix, wherein the first column is a vector of 1s for the constant factor, and the second column is 1762 observations of market returns represented by the Russell 3000 index, and

iv. E is a 1762×10 matrix of error terms.

The OLS estimation of the constant factor loading vector B in this multivariate regression is given by

$$\hat{B} = (F'F)^{-1}F'X,$$

which is essentially the same as the method used in a univariate simple linear regression. The total variance explained by the factor model $\hat{\Sigma}^{(F)}$ is given by

$$\hat{\Sigma}^{(F)} = cov(BM) + \hat{\Upsilon},$$

where $\hat{\Upsilon}$ is the estimated covariance matrix of the errors.

Based on the regression result, we produce the following table, wherein the first row shows the estimated constant factor loadings. The r^2's in the second row represent how much of the variance of the corresponding stock's return can be explained by the market factor M. The next 10 rows is the correlation matrix of stock return data, and the 10 rows in the middle is the correlation matrix of the factor model. The last 10 rows is the correlation matrix of the errors, and correlations with absolute values less than 0.1 are omitted. We expect the errors to be uncorrelated with each other, which is one of the key assumptions of factor models. The last correlation matrix shows that this factor model meets such expectation, as most of the errors are uncorrelated, suggesting that most of the systematic risk of these 10 stocks is captured by this factor model.

PCA Factor Model

Next, using the same stock return data, we perform a principal component analysis. Unlike the previous example, in this practice, we no longer benchmark individual stocks to the market factor. Instead, by rotating the data, we will have 10 principal components, which each is a linear combination of the 10 stocks. Then from these 10 principal components, we will choose a proper number of them to be our common factors. Comparison to macroeconomic factor models, apparently PCA factor models are difficult to interpret intuitively.

To determine the number of factors, we first consider how many principal components can explain a substantial proportion of the variance. Figure 1.2 shows the relationship between the proportion of variance explain (PoVE) and the principal components, and we can see that the first 6 components can roughly explain 90% of the variance, thus, we use the first 6 principal components as our common factors.

By using the first 6 principal components as common factors, we treat the remaining 4

Table 1.4: Key results from macroeconomic factor model

	AMZN	CPB	NVDA	CSCO	UNP	F	AIG	IBM	WMT	JCP
$\hat{B}$	1.07	0.5	1.35	0.99	1.08	1.27	1.4	0.78	0.47	1.07
r^2	0.28	0.22	0.32	0.37	0.54	0.49	0.43	0.43	0.21	0.12
AMZN	1	0.21	0.32	0.33	0.34	0.39	0.34	0.35	0.2	0.15
CPB	0.21	1	0.2	0.27	0.32	0.27	0.28	0.29	0.34	0.15
NVDA	0.32	0.2	1	0.41	0.38	0.42	0.35	0.38	0.18	0.19
CSCO	0.33	0.27	0.41	1	0.43	0.44	0.38	0.46	0.3	0.2
UNP	0.34	0.32	0.38	0.43	1	0.54	0.48	0.48	0.31	0.26
F	0.39	0.27	0.42	0.44	0.54	1	0.48	0.45	0.32	0.27
AIG	0.34	0.28	0.35	0.38	0.48	0.48	1	0.39	0.25	0.22
IBM	0.35	0.29	0.38	0.46	0.48	0.45	0.39	1	0.31	0.2
WMT	0.2	0.34	0.18	0.3	0.31	0.32	0.25	0.31	1	0.21
JCP	0.15	0.15	0.19	0.2	0.26	0.27	0.22	0.2	0.21	1
AMZN	1	0.21	0.32	0.33	0.35	0.39	0.35	0.35	0.2	0.15
CPB	0.21	1	0.2	0.27	0.32	0.27	0.28	0.29	0.34	0.15
NVDA	0.32	0.2	1	0.41	0.38	0.42	0.35	0.38	0.18	0.19
CSCO	0.33	0.27	0.41	1	0.43	0.44	0.38	0.46	0.3	0.2
UNP	0.35	0.32	0.38	0.43	1	0.54	0.48	0.48	0.31	0.26
F	0.39	0.27	0.42	0.44	0.54	1	0.48	0.45	0.32	0.27
AIG	0.35	0.28	0.35	0.38	0.48	0.48	1	0.39	0.25	0.22
IBM	0.35	0.29	0.38	0.46	0.48	0.45	0.39	1	0.31	0.2
WMT	0.2	0.34	0.18	0.3	0.31	0.32	0.25	0.31	1	0.21
JCP	0.15	0.15	0.19	0.2	0.26	0.27	0.22	0.2	0.21	1
AMZN	1									
CPB		1							0.16	
NVDA			1	0.1						
CSCO			0.1	1				0.1		
UNP					1					
F						1				
AIG							1			
IBM				0.1				1		
WMT		0.16							1	
JCP										1

principal components as errors, and we produce the following results as shown in table 1.5.

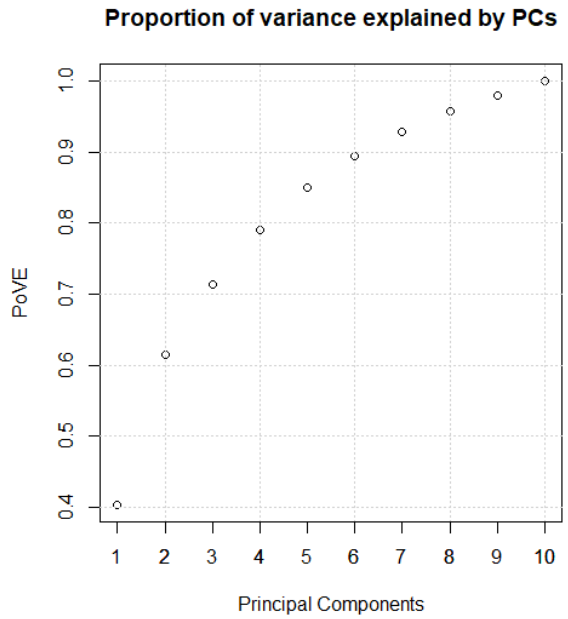

Figure 1.2: Proportion of variance explained the principal components

The first 6 rows are constant factor loadings of the first 6 principal components, which are essentially the first 6 eigenvectors from the rotation matrix Γ. The next row are the r^2's from the multivariate regression, indicating the proportion of the variance of a single stock explained by these 6 principal components. Compared to the previous macroeconomic factor model, this PCA model is able to explain much more variance of some stocks. The next 10 rows is the correlation matrix implied by factor model, and the last 10 rows is the correlation matrix of the errors, and correlations with absolute values less than 0.1 are omitted. As we see, this model does not produce uncorrelated errors, suggesting that it is not a good representation of the data.

Table 1.5: Key results from PCA model

	AMZN	CPB	NVDA	CSCO	UNP	F	AIG	IBM	WMT	JCP
$\hat{B}(PC1)$	0.28	0.1	0.4	0.25	0.25	0.33	0.36	0.18	0.11	0.59
$\hat{B}(PC2)$	-0.25	-0.05	-0.36	-0.18	-0.13	-0.18	-0.25	-0.11	-0.03	0.81
$\hat{B}(PC3)$	-0.29	-0.07	0.81	-0.02	-0.13	-0.15	-0.45	-0.05	-0.07	0.05
$\hat{B}(PC4)$	-0.84	0.04	0.02	0.03	0.1	0.07	0.52	0.01	0.02	-0.04
$\hat{B}(PC5)$	0.25	-0.11	0.22	-0.5	-0.25	-0.42	0.55	-0.21	-0.17	0.06
$\hat{B}(PC6)$	-0.03	-0.08	0.04	-0.69	0.13	0.68	-0.12	-0.1	-0.06	-0.02
r^2	1.00	0.21	1.00	0.93	0.58	0.93	1.00	0.47	0.25	1.00
AMZN	1	0.21	0.32	0.33	0.34	0.39	0.34	0.35	0.2	0.15
CPB	0.21	1	0.2	0.27	0.32	0.27	0.28	0.29	0.34	0.15
NVDA	0.32	0.2	1	0.41	0.38	0.42	0.35	0.38	0.18	0.19
CSCO	0.33	0.27	0.41	1	0.43	0.44	0.38	0.46	0.3	0.2
UNP	0.34	0.32	0.38	0.43	1	0.54	0.48	0.48	0.31	0.26
F	0.39	0.27	0.42	0.44	0.54	1	0.48	0.45	0.32	0.27
AIG	0.34	0.28	0.35	0.38	0.48	0.48	1	0.39	0.25	0.22
IBM	0.35	0.29	0.38	0.46	0.48	0.45	0.39	1	0.31	0.2
WMT	0.2	0.34	0.18	0.3	0.31	0.32	0.25	0.31	1	0.21
JCP	0.15	0.15	0.19	0.2	0.26	0.27	0.22	0.2	0.21	1
AMZN	1	-0.18	0.12	0.64	0.16	0.26	0.21	-0.97		0.13
CPB	-0.18	1	0.6	-0.48		-0.22	-0.46		0.15	-0.44
NVDA	0.12	0.6	1	-0.51	0.12	-0.36	-0.39	-0.26	0.83	-0.89
CSCO	0.64	-0.48	-0.51	1	-0.44	0.83	0.8	-0.6	-0.45	0.79
UNP	0.16		0.12	-0.44	1	-0.86	-0.84			-0.4
F	0.26	-0.22	-0.36	0.83	-0.86	1	0.95	-0.32	-0.23	0.7
AIG	0.21	-0.46	-0.39	0.8	-0.84	0.95	1	-0.23		0.64
IBM	-0.97		-0.26	-0.6		-0.32	-0.23	1		
WMT		0.15	0.83	-0.45		-0.23			1	-0.82
JCP	0.13	-0.44	-0.89	0.79	-0.4	0.7	0.64		-0.82	1

Part III

Appendix

1 code for Algorithm 1.3.2

1. Generate 3 covariance matrices and let $\mu = (1, 2)'$.

```
Sigma1<-matrix(c(1,-0.7,-0.7,1),2,2)
Sigma2<-matrix(c(1,0.7,0.7,1),2,2)
Sigma3<-matrix(c(1,0.3,0.3,1),2,2)
mu<-c(1,2)
```

2. Cholesky decomposition

```
A1<-chol(Sigma1)
A2<-chol(Sigma2)
A3<-chol(Sigma3)
```

3. Generate Z's.

```
n=5000
Z=matrix(rnorm(2*n),2,n)
```

4. Get X's.

```
X1 = matrix(rep(mu,10),2,n)+A1%*%Z
X2 = matrix(rep(mu,10),2,n)+A2%*%Z
X3 = matrix(rep(mu,10),2,n)+A3%*%Z
```

5. Visualization

```
library(MASS)
library(mvtnorm)
par(mfcol=c(3,2))
x.points <- seq(-3,3,length.out=100)
y.points <- x.points
z <- matrix(0,nrow=100,ncol=100)
```

```
kde1<-kde2d(X1[1,],X1[2,],n = 50)
image(kde1,xlab = 'x1', ylab = 'x2')
contour(kde1, add = TRUE)
title('simulation with corr = -0.7')

kde2<-kde2d(X2[1,],X2[2,],n = 50)
image(kde2,xlab = 'x1', ylab = 'x2')
contour(kde2, add = TRUE)
title('simulation with corr = 0.7')

kde3<-kde2d(X3[1,],X3[2,],n = 50)
image(kde3,xlab = 'x1', ylab = 'x2')
contour(kde3, add = TRUE)
title('simulation with corr = 0.3')

for (i in 1:100) { for (j in 1:100)
  { z[i,j] <- dmvnorm(c(x.points[i],y.points[j]), mean=mu,sigma=Sigma1) } }
contour(x.points,y.points,z,
        xlim = c(-2,4),ylim = c(-2,4),xlab = 'x1', ylab = 'x2')
title('real distribution with corr = -0.7')

for (i in 1:100) { for (j in 1:100)
  { z[i,j] <- dmvnorm(c(x.points[i],y.points[j]), mean=mu,sigma=Sigma2) } }
contour(x.points,y.points,z,
```

```
        xlim = c(-2,4),ylim = c(-2,4),xlab = 'x1', ylab = 'x2')
title('real distribution with corr = 0.7')


for (i in 1:100) { for (j in 1:100)
  { z[i,j] <- dmvnorm(c(x.points[i],y.points[j]), mean=mu,sigma=Sigma3) } }
contour(x.points,y.points,z,
        xlim = c(-2,4),ylim = c(-2,4),xlab = 'x1', ylab = 'x2')
title('real distribution with corr = 0.3')
```

2 code for Example 1.3

```
library(MVN)
```

```
## sROC 0.1-2 loaded
```

```
daily_rtn <- read.csv("path/to/file.csv")
monthly_rtn <- read.csv("path/to/file.csv")
quarterly_rtn <- read.csv("path/to/file.csv")
weekly_rtn <- read.csv("path/to/file.csv")
par(mfcol=c(2,2))
dailymardia<-mvn(daily_rtn[,-1],mvnTest = "mardia",
                 multivariatePlot = "qq")
legend(30, 10, legend=c("daily log returns"))
weeklymardia<-mvn(weekly_rtn[,-1],mvnTest = "mardia",
                  multivariatePlot = "qq")
legend(11, 10, legend=c("weekly log returns"))
```

```
monthlymardia<-mvn(monthly_rtn[,-1],mvnTest = "mardia",
                   multivariatePlot = "qq")
legend(10, 10, legend=c("monthly log returns"))
quarterlymardia<-mvn(quarterly_rtn[,-1],mvnTest = "mardia",
                     multivariatePlot = "qq")
legend(1, 23, legend=c("quarterly log returns"))
```

```
Table1.1<-rbind(dailymardia$multivariateNormality,
                weeklymardia$multivariateNormality,
               monthlymardia$multivariateNormality,
               quarterlymardia$multivariateNormality)
Table1.1<-Table1.1[Table1.1['Test']!='MVN',]
Type<-c('daily','daily','weekly','weekly',
        'monthly','monthly','quarterly','quarterly')
n<-c(dim(daily_rtn)[1],dim(daily_rtn)[1],dim(weekly_rtn)[1],
     dim(weekly_rtn)[1], dim(monthly_rtn)[1],dim(monthly_rtn)[1],
     dim(quarterly_rtn)[1],dim(quarterly_rtn)[1])
p<-c(0,0,0,0,0.021,0.001,0.06,0.717)
Table1.1[,3]<-p
Table1.1<-cbind(Type,n,Table1.1)
print(Table1.1,row.names = FALSE)
```