

# 6 Multivariate models

- 6.1 Basics of multivariate modelling
- 6.2 Normal mixture distributions
- 6.3 Spherical and elliptical distributions
- 6.4 Dimension reduction techniques

## 6.1 Basics of multivariate modelling

### 6.1.1 Random vectors and their distributions

#### Joint and marginal distributions

- Let  $\mathbf{X} = (X_1, \dots, X_d) : \Omega \rightarrow \mathbb{R}^d$  be a  $d$ -dimensional *random vector* (representing risk-factor changes, risks, etc.).
- The *(joint) distribution function (df)  $F$  of  $\mathbf{X}$*  is

$$F(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leq \mathbf{x}) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d), \quad \mathbf{x} \in \mathbb{R}^d.$$

- The  *$j$ th margin or marginal df  $F_j$  of  $\mathbf{X}$*  is

$$\begin{aligned} F_j(x_j) &= \mathbb{P}(X_j \leq x_j) \\ &= \mathbb{P}(X_1 \leq \infty, \dots, X_{j-1} \leq \infty, X_j \leq x_j, X_{j+1} \leq \infty, \dots, X_d \leq \infty) \\ &= F(\infty, \dots, \infty, x_j, \infty, \dots, \infty), \quad x_j \in \mathbb{R}, \quad j \in \{1, \dots, d\}. \end{aligned}$$

(interpreted as a *limit*).

- Similarly for *k-dimensional margins*. Suppose we partition  $\mathbf{X}$  into  $(\mathbf{X}'_1, \mathbf{X}'_2)'$ , where  $\mathbf{X}_1 = (X_1, \dots, X_k)'$  and  $\mathbf{X}_2 = (X_{k+1}, \dots, X_d)'$ , then the marginal distribution function of  $\mathbf{X}_1$  is

$$F_{\mathbf{X}_1}(\mathbf{x}_1) = \mathbb{P}(\mathbf{X}_1 \leq \mathbf{x}_1) = F(x_1, \dots, x_k, \infty, \dots, \infty).$$

- *F is absolutely continuous* if

$$F(\mathbf{x}) \underset{(*)}{=} \int_{-\infty}^{x_d} \dots \int_{-\infty}^{x_1} f(z_1, \dots, z_d) dz_1 \dots dz_d = \int_{(-\infty, \mathbf{x}]} f(\mathbf{z}) d\mathbf{z}$$

for some  $f \geq 0$  known as the *(joint) density of X (or F)*. Similarly, the *jth marginal df  $F_j$  is absolutely continuous* if  $F_j(x) = \int_{-\infty}^x f_j(z) dz$  for some  $f_j \geq 0$  known as the *density of  $X_j$  (or  $F_j$ )*.

- In case  $f$  exists,  $F_j(x_j) \underset{(*)}{=} \int_{-\infty}^{x_j} \int_{(-\infty, \infty)} f(\mathbf{z}) d\mathbf{z}_{-j} dz_j = \int_{-\infty}^{x_j} f_j(z_j) dz_j$ , so that  *$f_j(x_j)$  can be recovered from  $f$  via*

$$\underbrace{\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty}}_{d-1\text{-many}} f(z_1, \dots, z_{j-1}, x_j, z_{j+1}, \dots, z_d) dz_1 \dots dz_{j-1} dz_{j+1} \dots dz_d.$$

- Existence of a **joint density**  $\Rightarrow$  Existence of **marginal densities** for all  $k$ -dimensional marginals,  $1 \leq k \leq d-1$ . The **converse is false in general** (counter-examples can be constructed with **copulas**; see Chapter 7).
- By **replacing integrals by sums**, one obtains similar formulas for the **discrete case**, in which the notion of densities is replaced by **probability mass functions**.
- We sometimes work with the **survival function  $\bar{F}$  of  $\mathbf{X}$** ,

$$\bar{F}(\mathbf{x}) = \bar{F}_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} > \mathbf{x}) = \mathbb{P}(X_1 > x_1, \dots, X_d > x_d), \quad \mathbf{x} \in \mathbb{R}^d,$$

with corresponding  **$j$ th marginal survival function  $\bar{F}_j$**

$$\begin{aligned} \bar{F}_j(x_j) &= \mathbb{P}(X_j > x_j) \\ &= \bar{F}(-\infty, \dots, -\infty, x_j, -\infty, \dots, -\infty), \quad x_j \in \mathbb{R}, \quad j \in \{1, \dots, d\}. \end{aligned}$$

- Note that  $\bar{F}(\mathbf{x}) \neq 1 - F(\mathbf{x})$  in general (unless  $d = 1$ ).

## Conditional distributions and independence

- A **multivariate model** for risks in the form of a joint df, survival function or density, **implicitly describes** their **dependence structure**. We can then make statements about **conditional probabilities**.
- As before, consider  $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2) \sim F$ . The **conditional df of  $\mathbf{X}_2$  given  $\mathbf{X}_1 = \mathbf{x}_1$**  is  $F_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 | \mathbf{x}_1) = \mathbb{P}(\mathbf{X}_2 \leq \mathbf{x}_2 | \mathbf{X}_1 = \mathbf{x}_1) = \mathbb{E}(I_{\{\mathbf{X}_2 \leq \mathbf{x}_2\}} | \mathbf{X}_1 = \mathbf{x}_1)$ , where  $\mathbb{E}(\cdot | \cdot)$  denotes conditional expectation (**not discussed here**).
- A **useful identity** for conditional dfs is

$$F(\mathbf{x}) = \int_{(-\infty, \mathbf{x}_1]} F_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 | \mathbf{z}) dF_{\mathbf{X}_1}(\mathbf{z}); \quad (14)$$

see the appendix for a proof.

- ▶ If  $\mathbf{x}_1 \rightarrow \infty$ , then  $F_{\mathbf{X}_2}(\mathbf{x}_2) = \int_{\mathbb{R}^d} F_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 | \mathbf{z}) dF_{\mathbf{X}_1}(\mathbf{z})$ .
- ▶ If  $F$  has a density  $f$ , then  $f_{\mathbf{X}_2}(\mathbf{x}_2) = \int_{\mathbb{R}^d} f_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2 | \mathbf{z}) dF_{\mathbf{X}_1}(\mathbf{z})$ .

- If  $F$  has density  $f$  and  $f_{X_1}$  denotes the density of  $X_1$ , then

$$\begin{aligned} f(\mathbf{x}_1, \mathbf{x}_2) &= \frac{\partial^2}{\partial \mathbf{x}_2 \partial \mathbf{x}_1} F(\mathbf{x}_1, \mathbf{x}_2) \stackrel{(14)}{=} \frac{\partial}{\partial \mathbf{x}_2} F_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1) f_{X_1}(\mathbf{x}_1) \\ &= f_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1) f_{X_1}(\mathbf{x}_1). \end{aligned}$$

We call

$$f_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1) = \frac{f(\mathbf{x}_1, \mathbf{x}_2)}{f_{X_1}(\mathbf{x}_1)}$$

the *conditional density of  $X_2$  given  $X_1 = \mathbf{x}_1$* . In this case, the conditional df  $F_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1)$  is given by

$$F_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1) = \int_{-\infty}^{x_{k+1}} \cdots \int_{-\infty}^{x_d} f_{X_2|X_1}(z_{k+1}, \dots, z_d | \mathbf{x}_1) dz_{k+1} \cdots dz_d.$$

- $X_1, X_2$  are *independent* if  $F(\mathbf{x}_1, \mathbf{x}_2) = F_{X_1}(\mathbf{x}_1)F_{X_2}(\mathbf{x}_2)$  for all  $\mathbf{x}_1, \mathbf{x}_2$ .
- If  $F$  has density  $f$ , then  $X_1, X_2$  are independent if  $f(\mathbf{x}_1, \mathbf{x}_2) = f_{X_1}(\mathbf{x}_1)f_{X_2}(\mathbf{x}_2)$  for all  $\mathbf{x}_1, \mathbf{x}_2$ . In this case,  $f_{X_2|X_1}(\mathbf{x}_2 | \mathbf{x}_1) = f_{X_2}(\mathbf{x}_2)$ .

- The components  $X_1, \dots, X_d$  of  $\mathbf{X}$  are *(mutually) independent* if  $F(\mathbf{x}) = \prod_{j=1}^d F_j(x_j)$  for all  $\mathbf{x}$  or, if  $F$  has density  $f$ , if  $f(\mathbf{x}) = \prod_{j=1}^d f_j(x_j)$  for all  $\mathbf{x}$ .

## Moments and characteristic function

- If  $\mathbb{E}|X_j| < \infty$ ,  $j \in \{1, \dots, d\}$ , the *mean vector of  $\mathbf{X}$*  is defined by

$$\mathbb{E}\mathbf{X} = (\mathbb{E}X_1, \dots, \mathbb{E}X_d).$$

One can show:  $X_1, \dots, X_d$  independent  $\Rightarrow \mathbb{E}(X_1 \cdots X_d) = \prod_{j=1}^d \mathbb{E}(X_j)$

- If  $\mathbb{E}(X_j^2) < \infty$  for all  $j$ , the *covariance matrix of  $\mathbf{X}$*  is defined by

$$\text{cov}(\mathbf{X}) = \mathbb{E}((\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{X} - \mathbb{E}\mathbf{X})').$$

If we write  $\Sigma = \text{cov}(\mathbf{X})$ , its  $(i, j)$ th element is

$$\begin{aligned}\sigma_{ij} &= \Sigma_{ij} = \text{cov}(X_i, X_j) = \mathbb{E}((X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)) \\ &= \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j);\end{aligned}$$

the diagonal elements are  $\sigma_{ij} = \text{var}(X_j)$ ,  $j \in \{1, \dots, d\}$ .

- $X_1, X_2$  independent  $\not\Rightarrow \text{cov}(X_1, X_2) = 0$  (counter-examples can be constructed with **copulas**; see Chapter 7).
- The **cross covariance matrix between** two random vectors  $\mathbf{X}, \mathbf{Y}$  is defined by  $\text{cov}(\mathbf{X}, \mathbf{Y}) = \mathbb{E}((\mathbf{X} - \mathbb{E}\mathbf{X})(\mathbf{Y} - \mathbb{E}\mathbf{Y})')$ ; note that  $\text{cov}(\mathbf{X}, \mathbf{X}) = \text{cov}(\mathbf{X})$ .
- If  $\mathbb{E}(X_j^2) < \infty$ ,  $j \in \{1, \dots, d\}$ , the **correlation matrix of  $\mathbf{X}$**  is defined by the matrix **corr( $\mathbf{X}$ )** with  $(i, j)$ th element

$$\text{corr}(X_i, X_j) = \frac{\text{cov}(X_i, X_j)}{\sqrt{\text{var}(X_i) \text{var}(X_j)}}, \quad i, j \in \{1, \dots, d\},$$

which is in  $[-1, 1]$  with  $\text{corr}(X_i, X_j) = \pm 1$  if and only if  $X_j \stackrel{\text{a.s.}}{=} aX_i + b$  for some  $a \geq 0$  and  $b \in \mathbb{R}$ .

- **Some properties of  $\mathbb{E}()$  and  $\text{cov}()$ :**

1) For all  $A \in \mathbb{R}^{k \times d}$ ,  $\mathbf{b} \in \mathbb{R}^k$ :

$$\blacktriangleright \mathbb{E}(A\mathbf{X} + \mathbf{b}) = A\mathbb{E}\mathbf{X} + \mathbf{b} = A\boldsymbol{\mu} + \mathbf{b};$$



$$\begin{aligned} \blacktriangleright \text{cov}(A\mathbf{X} + \mathbf{b}) &= A \text{cov}(\mathbf{X}) A' = A\Sigma A'; \text{ if } k = 1 \ (A = \mathbf{a}'), \\ \mathbf{a}'\Sigma\mathbf{a} &= \text{cov}(\mathbf{a}'\mathbf{X}) = \text{var}(\mathbf{a}'\mathbf{X}) \geq 0, \quad \mathbf{a} \in \mathbb{R}^d, \end{aligned} \quad (15)$$

i.e. *covariance matrices are positive semidefinite*.

$$\blacktriangleright \text{cov}(\mathbf{X}_1 + \mathbf{X}_2) = \text{cov}(\mathbf{X}_1) + \text{cov}(\mathbf{X}_2) + 2 \text{cov}(\mathbf{X}_1, \mathbf{X}_2)$$

2) If  $\Sigma$  is a *positive definite matrix* (i.e.  $\mathbf{a}'\Sigma\mathbf{a} > 0$  for all  $\mathbf{a} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$ ), one can show that  $\Sigma$  is invertible.

3) A *symmetric, positive (semi)definite*  $\Sigma$  can be written as

$$\Sigma = AA' \quad \text{Cholesky decomposition} \quad (16)$$

for a lower triangular matrix  $A$  with  $A_{jj} > 0$  ( $A_{jj} \geq 0$ ) for all  $j$ .  $A$  is known as *Cholesky factor* (and also denoted by  $\Sigma^{1/2}$ ).

- Properties of  $\mathbf{X}$  can often be shown with the *characteristic function (cf)*

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(\exp(i\mathbf{t}'\mathbf{X})), \quad \mathbf{t} \in \mathbb{R}^d.$$

$X_1, \dots, X_d$  are independent  $\Leftrightarrow \phi_{\mathbf{X}}(\mathbf{t}) = \prod_{j=1}^d \phi_{X_j}(t_j)$  for all  $\mathbf{t}$ .

### Proposition 6.1 (Characterization of covariance matrices)

A symmetric matrix  $\Sigma$  is a covariance matrix if and only if it is positive semidefinite.

*Proof.*

“ $\Rightarrow$ ” As we have seen in (15), a covariance matrix  $\Sigma$  is positive semidefinite.

“ $\Leftarrow$ ” Let  $\Sigma$  be positive semidefinite with Cholesky factor  $A$ . Let  $\mathbf{X}$  be a random vector with  $\text{cov } \mathbf{X} = I_d = \text{diag}(1, \dots, 1)$  (e.g.  $X_j \stackrel{\text{ind.}}{\sim} N(0, 1)$ ). Then  $\text{cov}(A\mathbf{X}) = A \text{cov}(\mathbf{X}) A' = AA' = \Sigma$ , i.e.  $\Sigma$  is a covariance matrix (namely that of  $A\mathbf{X}$ ).  $\square$

#### 6.1.2 Standard estimators of covariance and correlation

- Assume  $\mathbf{X}_1, \dots, \mathbf{X}_n \sim F$  (daily/weekly/monthly/yearly risk-factor changes) to be **serially uncorrelated** (i.e. multivariate white noise) with  $\mu := \mathbb{E}\mathbf{X}_1$ ,  $\Sigma := \text{cov } \mathbf{X}_1$  and  $P = \text{corr}(\mathbf{X}_1)$ .

- Non-parametric **method-of-moments-like estimators** of  $\mu, \Sigma, P$  are

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \quad (\text{sample mean})$$

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' \quad (\text{sample covariance matrix})$$

$$\mathbf{R} = (R_{ij}) \text{ for } R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} \quad (\text{sample correlation matrix})$$

- Under joint normality ( $F$  multivariate normal),  $\bar{\mathbf{X}}, \mathbf{S}$  and  $\mathbf{R}$  are also **MLEs**.  $\mathbf{S}$  is biased, but an unbiased version can be obtained by

$$\mathbf{S}_n = \frac{n}{n-1} \mathbf{S}.$$

- Clearly,  $\bar{\mathbf{X}}$  is unbiased. Since the  $\mathbf{X}_i$ 's are uncorrelated,

$$\text{cov}(\bar{\mathbf{X}}) = \frac{1}{n^2} \sum_{i=1}^n \text{cov}(\mathbf{X}_i) = \frac{1}{n} \text{cov}(\mathbf{X}_1) = \frac{1}{n} \Sigma.$$

- $S_n$  is unbiased since

$$\begin{aligned}
 \mathbb{E}S_n &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}((\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})') \\
 &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}(((\mathbf{X}_i - \boldsymbol{\mu}) - (\bar{\mathbf{X}} - \boldsymbol{\mu}))((\mathbf{X}_i - \boldsymbol{\mu}) - (\bar{\mathbf{X}} - \boldsymbol{\mu}))') \\
 &= \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}((\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})' - (\bar{\mathbf{X}} - \boldsymbol{\mu})(\bar{\mathbf{X}} - \boldsymbol{\mu})') \\
 &= \frac{1}{n-1} \sum_{i=1}^n (\Sigma - \text{cov } \bar{\mathbf{X}}) \underset{\text{cov}(\bar{\mathbf{X}}) = \frac{\Sigma}{n}}{=} \frac{n}{n-1} \left(1 - \frac{1}{n}\right) \Sigma = \Sigma.
 \end{aligned}$$

- Further properties of  $\bar{\mathbf{X}}, S, R$  depend on  $F$ .

## 6.1.3 The multivariate normal distribution

### Definition 6.2 (Multivariate normal distribution)

$\mathbf{X} = (X_1, \dots, X_d)$  has a *multivariate normal* (or *Gaussian*) *distribution* if

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + A\mathbf{Z}, \quad (17)$$

where  $\mathbf{Z} = (Z_1, \dots, Z_k)$ ,  $Z_l \stackrel{\text{ind.}}{\sim} N(0, 1)$ ,  $A \in \mathbb{R}^{d \times k}$ ,  $\boldsymbol{\mu} \in \mathbb{R}^d$ .

- $\mathbb{E}\mathbf{X} = \boldsymbol{\mu} + A\mathbb{E}\mathbf{Z} = \boldsymbol{\mu}$
- $\text{cov}(\mathbf{X}) = \text{cov}(\boldsymbol{\mu} + A\mathbf{Z}) = A \text{cov}(\mathbf{Z})A' = AA' =: \Sigma$

### Proposition 6.3 (Cf of the multivariate normal distribution)

Let  $\mathbf{X}$  be as in (17) and  $\Sigma = AA'$ . Then the cf of  $\mathbf{X}$  is

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(\exp(it'\mathbf{X})) = \exp\left(it'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right), \quad \mathbf{t} \in \mathbb{R}^d.$$

*Idea of proof.* Using the fact that  $\phi_Z(t) = \exp(-t^2/2)$  for  $Z \sim N(0, 1)$  (see the appendix for a proof) and letting  $\tilde{t}' = t' A$ , we obtain that

$$\begin{aligned}\phi_X(t) &= \mathbb{E}(\exp(it'(\mu + AZ))) = \exp(it'\mu)\mathbb{E}(\exp(i\tilde{t}'Z)) \\ &\stackrel{\text{ind.}}{=} \exp(it'\mu) \prod_{j=1}^d \mathbb{E}(\exp(i\tilde{t}_j Z_j)) = \exp\left(it'\mu - \frac{1}{2} \sum_{j=1}^d \tilde{t}_j^2\right) \\ &= \exp\left(it'\mu - \frac{1}{2} \tilde{t}'\tilde{t}\right) = \exp\left(it'\mu - \frac{1}{2} t' A A' t\right) \\ &= \exp\left(it'\mu - \frac{1}{2} t' \Sigma t\right) \quad \square\end{aligned}$$

- We see that the multivariate normal distribution is characterized by  $\mu$  and  $\Sigma$ , hence the notation  $X \sim N_d(\mu, \Sigma)$ .
- $N_d(\mu, \Sigma)$  can be characterized by univariate normal distributions.

### Proposition 6.4 (Characterization of $N_d(\boldsymbol{\mu}, \Sigma)$ )

$$\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma) \iff \mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a}) \quad \forall \mathbf{a} \in \mathbb{R}^d.$$

*Proof.* “ $\Rightarrow$ ” via uniqueness of cfs; “ $\Leftarrow$ ” via Corollary A.10 □

#### Consequences:

- Margins:  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma) \xRightarrow{\mathbf{a}=\mathbf{e}_j} X_j \sim N(\mu_j, \sigma_{jj}^2), \quad j \in \{1, \dots, d\}.$
- Sums:  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma) \xRightarrow{\mathbf{a}=\mathbf{1}} \sum_{j=1}^d X_j \sim N(\sum_{j=1}^d \mu_j, \sum_{i,j} \sigma_{ij}).$

### Proposition 6.5 (Density)

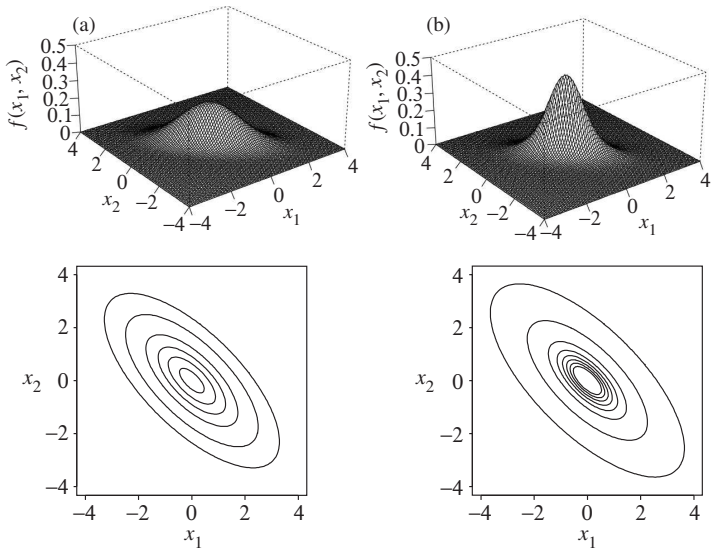
Let  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  with  $\text{rank } \Sigma = d = k \Rightarrow \Sigma$  pos. definite, invertible). Via the Density Transformation Theorem, it is an exercise to show that  $\mathbf{X}$  has density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad \mathbf{x} \in \mathbb{R}^d.$$

### Consequences:

- Sets of the form  $S_c = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c\}$ ,  $c > 0$ , describe points of equal density. Contours of equal density are thus ellipsoids. Whenever a multivariate density  $f_{\mathbf{X}}(\mathbf{x})$  depends on  $\mathbf{x}$  only through the quadratic form  $(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$ , it is the density of an elliptical distribution (see later).
- The components of  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  are mutually independent if and only if  $\Sigma$  is diagonal, i.e. if and only if the components of  $\mathbf{X}$  are uncorrelated.





Left:  $N_d(\boldsymbol{\mu}, \Sigma)$  for  $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ,  $\Sigma = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 1 \end{pmatrix}$ ; Right:  $t_\nu(\boldsymbol{\mu}, \frac{\nu-2}{\nu}\Sigma)$ ,  $\nu = 4$ ,  
 (same mean and covariance matrix as on the left-hand side)

The definition of  $N_d(\boldsymbol{\mu}, \Sigma)$  in terms of a **stochastic representation** ( $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + A\mathbf{Z}$ ) **directly justifies the following sampling algorithm.**

### **Algorithm 6.6 (Sampling $N_d(\boldsymbol{\mu}, \Sigma)$ )**

Let  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  with  $\Sigma$  positive definite.

- 1) Compute the Cholesky factor  $A$  of  $\Sigma$ ; see, e.g. Press et al. (1992).
- 2) Generate  $Z_j \stackrel{\text{ind.}}{\sim} N(0, 1)$ ,  $j \in \{1, \dots, d\}$ .
- 3) Return  $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Z}$ , where  $\mathbf{Z} = (Z_1, \dots, Z_d)$ .

## **Further useful properties of multivariate normal distributions**

### ■ **Linear combinations**

If  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  and  $B \in \mathbb{R}^{k \times d}$ ,  $\mathbf{b} \in \mathbb{R}^k$ , then

$$\begin{aligned} B\mathbf{X} + \mathbf{b} &= B(\boldsymbol{\mu} + A\mathbf{Z}) + \mathbf{b} = (B\boldsymbol{\mu} + \mathbf{b}) + B A \mathbf{Z} \\ &\sim N_k(B\boldsymbol{\mu} + \mathbf{b}, B A (B A)') = N_k(B\boldsymbol{\mu} + \mathbf{b}, B \Sigma B'). \end{aligned}$$

Special case (see variance-covariance method; or Proposition 6.4):  
 $\mathbf{b}'\mathbf{X} \sim N(\mathbf{b}'\boldsymbol{\mu}, \mathbf{b}'\Sigma\mathbf{b})$

### ■ Marginal dfs

Let  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  and write  $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2)$ , where  $\mathbf{X}_1 \in \mathbb{R}^k$ ,  $\mathbf{X}_2 \in \mathbb{R}^{d-k}$ , and  $\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \boldsymbol{\mu}'_2)$ ,  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ . Then

$$\mathbf{X}_1 \sim N_k(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \mathbf{X}_2 \sim N_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}).$$

*Proof.* Choose  $B = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$  and  $B = \begin{pmatrix} 0 & 0 \\ 0 & I_{d-k} \end{pmatrix}$ , respectively, in the above.

### ■ Conditional distributions

Let  $\mathbf{X}$  be as before and  $\Sigma$  be positive definite. One can show that

$$\mathbf{X}_2 \mid \mathbf{X}_1 = \mathbf{x}_1 \sim N_{d-k}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1}),$$

where  $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$  and  $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ .

### ■ Quadratic forms

Let  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  and  $\Sigma$  be positive definite with Cholesky factor  $A$ .

Furthermore, let  $\mathbf{Z} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})$ . Then  $\mathbf{Z} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$ . Moreover,

$$(\mathbf{X} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{Z}' \mathbf{Z} \sim \chi_d^2, \quad (18)$$

which is useful for (goodness-of-fit) testing of  $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ; see later.

## ■ Convolutions

Let  $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  and  $\mathbf{Y} \sim \mathcal{N}_d(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$  be independent. Via cfs it is then an exercise to show that

$$\mathbf{X} + \mathbf{Y} \sim \mathcal{N}_d(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \boldsymbol{\Sigma} + \tilde{\boldsymbol{\Sigma}}).$$

### 6.1.4 Testing multivariate normality

- For testing univariate normality, all tests of Section 3.1.2 can be applied.
- Now consider multivariate normality. By Proposition 6.4,

$$\mathbf{X}_1, \dots, \mathbf{X}_n \stackrel{\text{ind.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Rightarrow \mathbf{a}' \mathbf{X}_1, \dots, \mathbf{a}' \mathbf{X}_n \stackrel{\text{ind.}}{\sim} \mathcal{N}(\mathbf{a}' \boldsymbol{\mu}, \mathbf{a}' \boldsymbol{\Sigma} \mathbf{a}).$$

This can be tested statistically (for some  $\mathbf{a}$ ) with various goodness-of-fit tests (e.g. Q-Q plots) known for univariate normality (however, for

$\mathbf{a} = \mathbf{e}_j$ ,  $j \in \{1, \dots, d\}$ , we would **only test normality of the margins**, **not joint normality**). Alternatively, (18) can be used to test joint normality.

- Multivariate Shapiro–Wilk

- Mardia's test

- ▶ According to (18), if  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$  with  $\Sigma$  positive definite, then  $(\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$ .
- ▶ Let  $D_i^2 = (\mathbf{X}_i - \bar{\mathbf{X}})' S^{-1} (\mathbf{X}_i - \bar{\mathbf{X}})$  denote the **squared Mahalanobis distances** and  $D_{ij} = (\mathbf{X}_i - \bar{\mathbf{X}})' S^{-1} (\mathbf{X}_j - \bar{\mathbf{X}})$  the **Mahalanobis angles**.
- ▶ Let  $b_d = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n D_{ij}^3$  and  $k_d = \frac{1}{n} \sum_{i=1}^n D_i^4$ . **Under the null hypothesis** one can show that asymptotically for  $n \rightarrow \infty$ ,

$$\frac{n}{6} b_d \sim \chi_{d(d+1)(d+2)/6}^2, \quad \frac{k_d - d(d+2)}{\sqrt{8d(d+2)/n}} \sim N(0, 1),$$

which can be used for testing; see Joensuu and Vogel (2014).

### Example 6.7 (Multivariate (non-)normality of 10 Dow Jones stocks)

- We apply Mardia's test (of multivariate skewness and kurtosis) to daily/weekly/monthly/quarterly log-returns of 10 (of the 30) Dow Jones stocks from 1993–2000.

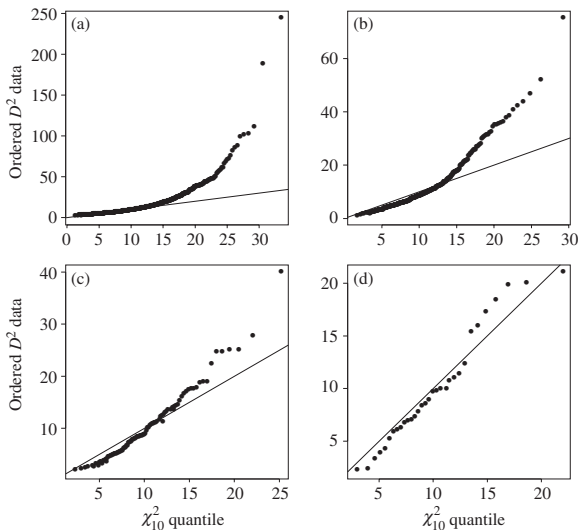
$n$	Daily 2020	Weekly 416	Monthly 96	Quarterly 32
$b_{10}$	9.31	9.91	21.10	50.10
$p$ -value	0.00	0.00	0.00	0.02
$k_{10}$	242.45	177.04	142.65	120.83
$p$ -value	0.00	0.00	0.00	0.44

- We also compare  $D_i^2$  data to a  $\chi_{10}^2$  using a Q-Q plot; see the next page.

**Conclusion:** Daily/weekly/monthly data: Evidence against joint normality; Quarterly data: CLT effect seems to take place (but too little data to say more); still evidence against joint normality.

Q-Q plot of  $D_i^2$  data against a  $\chi_{10}^2$  distribution:

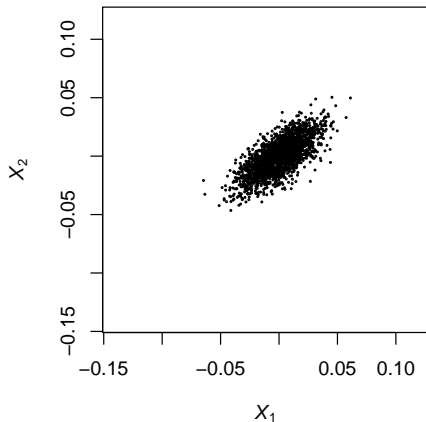
(a) daily data; (b) weekly data; (c) monthly data; and (d) quarterly data



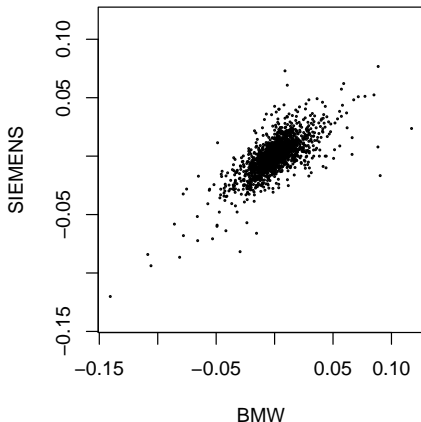
## Example 6.8 (Simulated data vs BMW–Siemens)

Is the **BMW–Siemens data** (see Section 3.2.2) **jointly normal**?

**Simulated data (fitted multivariate normal)**



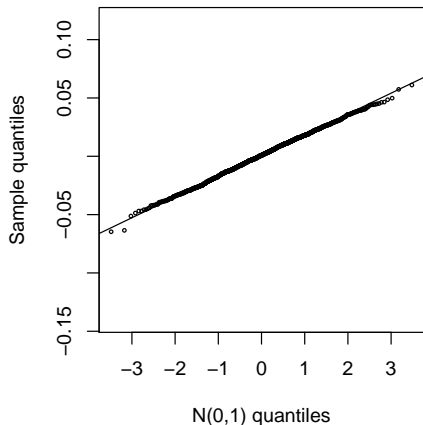
**Real risk-factor changes**



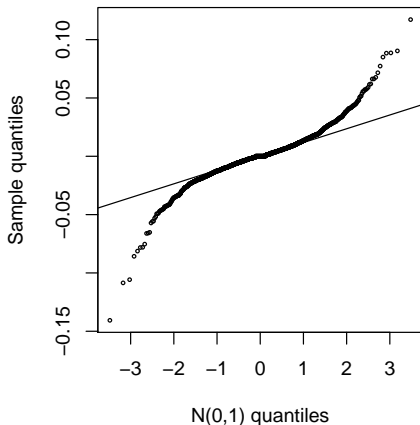


Considering the **first margin** only:

**Q-Q plot for margin 1 (simulated data)**

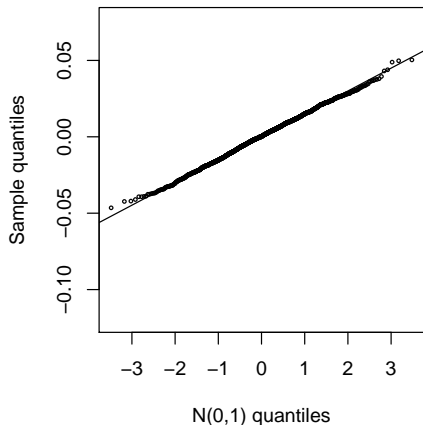


**Q-Q plot for margin 1 (real data)**

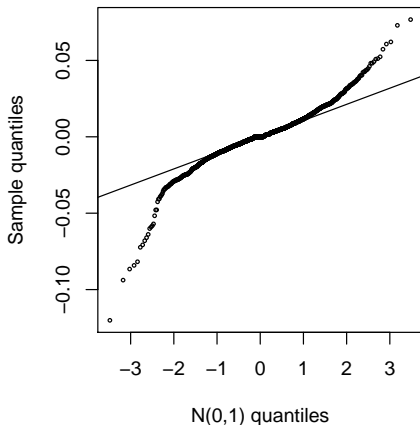


Considering the **second margin** only:

**Q-Q plot for margin 2 (simulated data)**

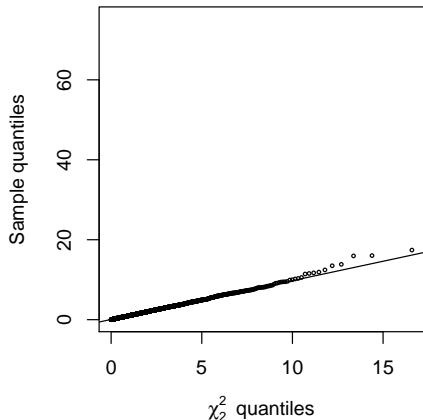


**Q-Q plot for margin 2 (real data)**

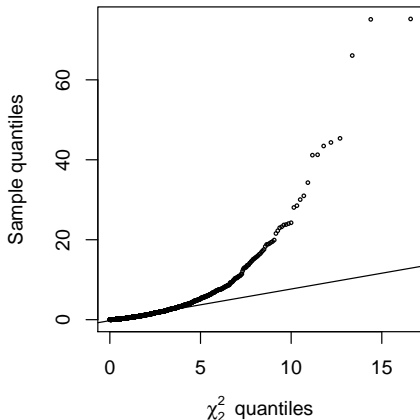


Q-Q plot of the simulated (left) or real (right)  $D_i^2$ 's against a  $\chi_2^2$ :

Q-Q plot of  $D_i^2$  (simulated data)



Q-Q plot of  $D_i^2$  (real data)



## Advantages of $N_d(\mu, \Sigma)$

- Inference “easy”.
- Distribution is determined by  $\mu$  and  $\Sigma$ .
- Linear combinations are normal ( $\Rightarrow$   $\text{VaR}_\alpha$  and  $\text{ES}_\alpha$  calculations for portfolios are easy).
- Marginal distributions are normal.
- Conditional distributions are normal.
- Quadratic forms are known.
- Convolutions are normal.
- Sampling is straightforward.
- Independence and uncorrelatedness are equivalent.

## Drawbacks of $N_d(\boldsymbol{\mu}, \Sigma)$ for modelling risk-factor changes

- 1) Tails of univariate (normal) margins are too thin (generate too few extreme events).
- 2) Joint tails are too thin (generate too few joint extreme events).  
 $N_d(\boldsymbol{\mu}, \Sigma)$  cannot capture the notion of tail dependence (see Chapter 7).
- 3) Very strong symmetry known as radial symmetry:  $\mathbf{X}$  is called *radially symmetric about  $\boldsymbol{\mu}$*  if  $\mathbf{X} - \boldsymbol{\mu} \stackrel{d}{=} \boldsymbol{\mu} - \mathbf{X}$ . This is true for  $N_d(\boldsymbol{\mu}, \Sigma)$ .

### In short:

- Normal variance mixture distributions can address 1) and 2) while sharing many of the desirable properties of  $N_d(\boldsymbol{\mu}, \Sigma)$ .
- Normal mean-variance mixture distribution can also address 3) (but at the expense of tractability in comparison to  $N_d(\boldsymbol{\mu}, \Sigma)$ ).

## 6.2 Normal mixture distributions

Idea: Randomize  $\Sigma$  (and  $\mu$ ) with a non-negative rv  $W$ .

### 6.2.1 Normal variance mixtures

#### Definition 6.9 (Multivariate normal variance mixtures)

The random vector  $\mathbf{X}$  has a (multivariate) *normal variance mixture distribution* if

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W} \mathbf{A} \mathbf{Z}, \quad (19)$$

where  $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$ ,  $W \geq 0$  is a rv independent of  $\mathbf{Z}$ ,  $\mathbf{A} \in \mathbb{R}^{d \times k}$ , and  $\boldsymbol{\mu} \in \mathbb{R}^d$ .  $\boldsymbol{\mu}$  is called *location vector* and  $\Sigma = \mathbf{A} \mathbf{A}'$  *scale* (or *dispersion*) *matrix*.

Observe that  $(\mathbf{X} | W = w) \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{w} \mathbf{A} \mathbf{Z} = N_d(\boldsymbol{\mu}, w \mathbf{A} \mathbf{A}') = N_d(\boldsymbol{\mu}, w \Sigma)$ ; or  $(\mathbf{X} | W) \stackrel{d}{=} N_d(\boldsymbol{\mu}, W \Sigma)$ .  $W$  can be interpreted as a *shock* affecting the variances of all risk factors.

## Properties of multivariate normal variance mixtures

Let  $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}A\mathbf{Z}$  and  $\mathbf{Y} = \boldsymbol{\mu} + A\mathbf{Z}$ . Assume that  $\text{rank}(A) = d \leq k$  and that  $\Sigma$  is positive definite.

- If  $\mathbb{E}\sqrt{W} < \infty$ , then  $\mathbb{E}(\mathbf{X}) \stackrel{\text{ind.}}{=} \boldsymbol{\mu} + \mathbb{E}(\sqrt{W})A\mathbb{E}(\mathbf{Z}) = \boldsymbol{\mu} + \mathbf{0} = \boldsymbol{\mu} = \mathbb{E}\mathbf{Y}$
- If  $\mathbb{E}W < \infty$ , then

$$\begin{aligned}\text{cov}(\mathbf{X}) &= \text{cov}(\sqrt{W}A\mathbf{Z}) = \mathbb{E}((\sqrt{W}A\mathbf{Z})(\sqrt{W}A\mathbf{Z})') \\ &\stackrel{\text{ind.}}{=} \mathbb{E}(W) \cdot \mathbb{E}(A\mathbf{Z}\mathbf{Z}'A') = \mathbb{E}(W) \cdot A\mathbb{E}(\mathbf{Z}\mathbf{Z}')A' \\ &= \mathbb{E}(W)AI_kA' = \mathbb{E}(W)\Sigma \neq \Sigma \quad (\text{in general } (= \text{cov}(\mathbf{Y})))\end{aligned}$$

- However, if they exist (i.e. if  $\mathbb{E}W < \infty$ ), it is easy to check that  $\text{corr}(\mathbf{X})$  and  $\text{corr}(\mathbf{Y})$  are equal.

### Lemma 6.10 (Independence in normal variance mixtures)

Let  $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{Z}$  with  $\mathbb{E}W < \infty$  (uncorrelated normal variance mixture). Then

$X_i$  and  $X_j$  are independent  $\iff W$  is a.s. constant (i.e.  $\mathbf{X} \sim \mathcal{N}_d$ ).

See the appendix for a proof. Intuitively,  $W$  affects all components of  $\mathbf{X}$  and thus creates dependence (unless it is constant).

**Recall:** If  $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$ , then  $\phi_{\mathbf{X}}(\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t})$ .

Furthermore,  $\mathbf{X} \mid W = w \sim \mathcal{N}_d(\boldsymbol{\mu}, w\Sigma)$

- **Characteristic function:** The cf of a multivariate normal variance mixtures is

$$\begin{aligned}\phi_{\mathbf{X}}(\mathbf{t}) &= \mathbb{E}(\exp(i\mathbf{t}'\mathbf{X})) = \mathbb{E}(\mathbb{E}(\exp(i\mathbf{t}'\mathbf{X}) \mid W)) \\ &= \mathbb{E}(\exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}W\mathbf{t}'\Sigma\mathbf{t})) = \exp(i\mathbf{t}'\boldsymbol{\mu})\mathbb{E}(\exp(-W\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t})).\end{aligned}$$



- **LS transform:** The *Laplace-Stieltjes transform* of  $F_W$  is

$$\hat{F}_W(\theta) := \mathbb{E}(\exp(-\theta W)) = \int_0^\infty e^{-\theta w} dF_W(w).$$

Therefore,  $\phi_{\mathbf{X}}(\mathbf{t}) = \exp(it'\boldsymbol{\mu})\hat{F}_W(\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t})$ . We thus introduce the notation  $\mathbf{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$  for a  $d$ -dimensional multivariate normal variance mixture.

- **Density:** If  $\Sigma$  is positive definite,  $\mathbb{P}(W = 0) = 0$ , the density of  $\mathbf{X}$  is

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \int_0^\infty f_{\mathbf{X}|W}(\mathbf{x} | w) dF_W(w) \\ &= \int_0^\infty \frac{1}{(2\pi)^{d/2} w^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}{2w}\right) dF_W(w). \end{aligned}$$

$\Rightarrow$  Only depends on  $\mathbf{x}$  through  $(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})$ .

$\Rightarrow$  Multivariate normal variance mixtures are **elliptical distributions**.

If  $\Sigma$  is diagonal and  $\mathbb{E}W < \infty$ ,  $\mathbf{X}$  is **uncorrelated** (as  $\text{cov}(\mathbf{X}) = \mathbb{E}(W)\Sigma$ )  
**but not independent** unless  $W$  is constant a.s.

- **Linear combinations:** For  $\mathbf{X} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$  and  $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$ , where  $\mathbf{B} \in \mathbb{R}^{k \times d}$  and  $\mathbf{b} \in \mathbb{R}^k$ , we have  $\mathbf{Y} \sim M_k(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}', \hat{F}_W)$ ; this can be shown via cfs. If  $\mathbf{a} \in \mathbb{R}^d$  ( $\mathbf{b} = \mathbf{0}$ ,  $\mathbf{B} = \mathbf{a}' \in \mathbb{R}^{1 \times d}$ ),  $\mathbf{a}'\mathbf{X} \sim M_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a}, \hat{F}_W)$ .
- **Sampling:**

**Algorithm 6.11 (Simulation of  $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z} \sim M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ )**

- 1) Generate  $\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$ .
- 2) Generate  $W \sim F_W$  (with LS transform  $\hat{F}_W$ ), independent of  $\mathbf{Z}$ .
- 3) Compute the Cholesky factor  $\mathbf{A}$  (such that  $\mathbf{A}\mathbf{A}' = \Sigma$ ).
- 4) Return  $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{A}\mathbf{Z}$ .

**Example 6.12 ( $t_d(\nu, \boldsymbol{\mu}, \Sigma)$  distribution)**

For Step 2), generate  $V \sim \chi_\nu^2$  and set  $W = \frac{\nu}{V} \sim \text{Ig}(\nu/2, \nu/2)$ . Or:  $W = \frac{1}{V}$  with  $V \sim \Gamma(\frac{\nu}{2}, \frac{\nu}{2})$  ( $\Gamma(\alpha, \beta)$  density:  $f(x) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$ ).

## Examples of multivariate normal variance mixtures

- **Multivariate normal distribution**

$W = 1$  a.s. (degenerate case)

- **Two point mixture**

$$W = \begin{cases} w_1 & \text{with probability } p, \\ w_2 & \text{with probability } 1 - p \end{cases} \quad w_1, w_2 > 0, w_1 \neq w_2.$$

Can be used to model **ordinary and stress regimes**; extends to  $k$  regimes.

- **Symmetric generalised hyperbolic distribution**

$W$  has a generalised inverse Gaussian distribution (GIG); see McNeil et al. (2015, p. 187)

- **Multivariate  $t$  distribution**

$W$  has an inverse gamma distribution  $W = 1/V$  for  $V \sim \Gamma(\nu/2, \nu/2)$ .

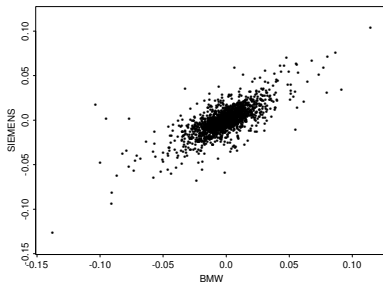
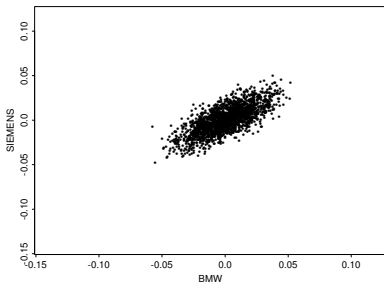
- ▶  $\mathbb{E}(W) = \frac{\nu}{\nu-2} \Rightarrow \text{cov}(\mathbf{X}) = \frac{\nu}{\nu-2} \Sigma$ . For finite variances/correlations,  $\nu > 2$  is required. For finite mean,  $\nu > 1$  is required.

- ▶ The density of the multivariate  $t$  distribution is given by

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)(\nu\pi)^{d/2}|\Sigma|^{1/2}} \left( 1 + \frac{(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\nu} \right)^{-\frac{\nu+d}{2}},$$

where  $\boldsymbol{\mu} \in \mathbb{R}^d$ ,  $\Sigma \in \mathbb{R}^{d \times d}$  is a positive definite matrix, and  $\nu$  is the degrees of freedom. Notation:  $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$ .

- ▶  $t_d(\nu, \boldsymbol{\mu}, \Sigma)$  has heavier marginal and joint tails than  $N_d(\boldsymbol{\mu}, \Sigma)$ .
- ▶ BMW–Siemens data: Simulations from fitted  $N_d(\boldsymbol{\mu}, \Sigma)$  and  $t_d(3, \boldsymbol{\mu}, \Sigma)$ :



## 6.2.2 Normal mean-variance mixtures

- Radial symmetry implies that **all one-dimensional margins of normal variance mixtures are symmetric**.
- Often visible in data: **joint losses have heavier tails** than joint gains.

**Idea:** Introduce **asymmetry by mixing** normal distributions **with different means and variances**.

$\mathbf{X}$  has a (multivariate) *normal mean-variance mixture distribution* if

$$\mathbf{X} \stackrel{d}{=} \mathbf{m}(W) + \sqrt{W} \mathbf{A} \mathbf{Z}, \quad (20)$$

where

- $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$ ;
- $W \geq 0$  is a scalar random variable which is independent of  $\mathbf{Z}$ ;
- $\mathbf{A} \in \mathbb{R}^{d \times k}$  is a matrix of constants;
- $\mathbf{m} : [0, \infty) \rightarrow \mathbb{R}^d$  is a measurable function.

- Normal mean-variance mixtures add **skewness**: Let  $\Sigma = AA'$  and observe that  $\mathbf{X} | W = w \sim N_d(\mathbf{m}(w), w\Sigma)$ . In general, **they are no longer elliptical** (see later).

### Example 6.13

- Suppose we have  $\mathbf{m}(W) = \boldsymbol{\mu} + W\boldsymbol{\gamma}$ . Since

$$\mathbb{E}(\mathbf{X} | W) = \boldsymbol{\mu} + W\boldsymbol{\gamma},$$

$$\text{cov}(\mathbf{X} | W) = W\Sigma$$

we have

$$\mathbb{E}\mathbf{X} = \mathbb{E}(\mathbb{E}(\mathbf{X} | W)) = \boldsymbol{\mu} + \mathbb{E}(W)\boldsymbol{\gamma} \quad \text{if } \mathbb{E}W < \infty,$$

$$\begin{aligned} \text{cov}(\mathbf{X}) &= \mathbb{E}(\text{cov}(\mathbf{X} | W)) + \text{cov}(\mathbb{E}(\mathbf{X} | W)) \\ &= \mathbb{E}(W)\Sigma + \text{var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}' \quad \text{if } \mathbb{E}(W^2) < \infty. \end{aligned}$$

- If  $W$  has a GIG distribution, then  $\mathbf{X}$  follows a *generalised hyperbolic distribution*.  $\boldsymbol{\gamma} = \mathbf{0}$  leads to (elliptical) normal variance mixtures; see McNeil et al. (2015, Sections 6.2.3) for details.

## 6.3 Spherical and elliptical distributions

Empirical examples (see McNeil et al. (2015, Sections 6.2.4)) show that

- 1)  $M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$  (e.g. multivariate  $t$ , NIG) provide superior models to  $N_d(\boldsymbol{\mu}, \Sigma)$  for daily/weekly US stock-return data;
- 2) the more general skewed normal mean-variance mixture distributions offer only a modest improvement.

We soon study elliptical distributions, a generalization of  $M_d(\boldsymbol{\mu}, \Sigma, \hat{F}_W)$ .

### 6.3.1 Spherical distributions

#### Definition 6.14 (Spherical distribution)

A random vector  $\mathbf{Y} = (Y_1, \dots, Y_d)$  has a spherical distribution if for every orthogonal  $U \in \mathbb{R}^{d \times d}$  (i.e.  $U \in \mathbb{R}^{d \times d}$  with  $UU' = U'U = I_d$ )

$$\mathbf{Y} \stackrel{d}{=} U\mathbf{Y} \quad (\text{distributionally invariant under rotations and reflections})$$

### Theorem 6.15 (Characterization of spherical distributions)

Let  $\|t\| = (t_1^2 + \dots + t_d^2)^{1/2}$ ,  $t \in \mathbb{R}^d$ . The following are equivalent:

- 1)  $Y$  is spherical (notation:  $Y \sim S_d(\psi)$  for  $\psi$  as below).
- 2)  $\exists$  a characteristic generator  $\psi : [0, \infty) \rightarrow \mathbb{R}$ , such that  $\phi_Y(t) = \mathbb{E}(e^{it'Y}) = \psi(\|t\|^2)$ ,  $\forall t \in \mathbb{R}^d$ .
- 3) For every  $a \in \mathbb{R}^d$ ,  $a'Y \stackrel{d}{=} \|a\|Y_1$  (lin. comb. are of the same type).  
 $\Rightarrow$  Subadditivity of  $\text{VaR}_\alpha$  for jointly elliptical losses

### Theorem 6.16 (Stochastic representation)

$Y \sim S_d(\psi)$  if and only if  $Y \stackrel{d}{=} RS$  for an independent radial part  $R \geq 0$  and  $S \sim U(\{x \in \mathbb{R}^d : \|x\| = 1\})$ .

- See the appendix for proofs for Theorems 6.15 and 6.16.
- If  $Y$  has a density  $f_Y$ , it satisfies  $f_Y(y) = g(\|y\|^2)$  for a function  $g : [0, \infty) \rightarrow [0, \infty)$  referred to as density generator (i.e.  $f_Y$  is constant on spheres); see the appendix for a proof.



### Corollary 6.17

If  $\mathbf{Y} \sim S_d(\psi)$  and  $\mathbb{P}(\mathbf{Y} = \mathbf{0}) = 0$ , then  $(\|\mathbf{Y}\|, \frac{\mathbf{Y}}{\|\mathbf{Y}\|}) \stackrel{d}{=} (R, \mathbf{S})$  since

$$(\|\mathbf{Y}\|, \frac{\mathbf{Y}}{\|\mathbf{Y}\|}) \stackrel{d}{=} (\|R\mathbf{S}\|, \frac{R\mathbf{S}}{\|R\mathbf{S}\|}) = (|R|\|\mathbf{S}\|, \frac{\mathbf{S}}{\|\mathbf{S}\|}) = (R, \mathbf{S}).$$

In particular,  $\|\mathbf{Y}\|$  and  $\mathbf{Y}/\|\mathbf{Y}\|$  are independent ( $\Rightarrow$  goodness-of-fit).

### Example 6.18 (Standardized multivariate normal variance mixtures)

- $\mathbf{Y} \sim M_d(\mathbf{0}, I_d, \hat{F}_W)$  is spherical (recall:  $\mathbf{Y} \stackrel{d}{=} \mathbf{0} + \sqrt{W}I_d\mathbf{Z}$ ) since

$$\begin{aligned}\phi_{\mathbf{Y}}(\mathbf{t}) &= \mathbb{E}(\exp(i\mathbf{t}'\sqrt{W}\mathbf{Z})) = \mathbb{E}_W(\mathbb{E}(\exp(i\mathbf{t}\sqrt{W})'\mathbf{Z}) \mid W)) \\ &= \mathbb{E}(\exp(-\tfrac{1}{2}W\mathbf{t}'\mathbf{t})) = \hat{F}_W(\tfrac{1}{2}\mathbf{t}'\mathbf{t}) = \hat{F}_W(\tfrac{1}{2}\|\mathbf{t}\|^2),\end{aligned}$$

so  $\mathbf{Y} \sim S_d(\psi)$  by Theorem 6.15 Part 2). We thus have  $\psi(t) = \hat{F}_W(t/2)$ .

- For  $\mathbf{Y} \sim N_d(\mathbf{0}, I_d)$ ,  $\psi(t) = \exp(-t/2)$ . By Corollary 6.17, simulating  $\mathbf{S} \sim U(\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| = 1\})$  can thus be done via  $\mathbf{S} \stackrel{d}{=} \mathbf{Y}/\|\mathbf{Y}\|$ . Fang et al. (1990, pp. 48) show that  $\psi$  generates  $S_d(\psi)$  for all  $d \in \mathbb{N}$  if and only if it is the characteristic generator of a normal mixture.

### Example 6.19 ( $R$ , $S$ , $\text{cov}$ , $\text{corr}$ )

- It follows from  $\mathbf{Y} \sim N_d(\mathbf{0}, I_d)$  and  $R^2 = \mathbf{Y}'\mathbf{Y} \sim \chi_d^2$  that

$$\mathbf{0} = \mathbb{E}\mathbf{Y} \underset{\text{Th. 6.16}}{=} \mathbb{E}R\mathbb{E}\mathbf{S} \Rightarrow \mathbb{E}\mathbf{S} = \mathbf{0},$$

$$I_d = \underset{\text{Th. 6.16}}{\text{cov}} \mathbf{Y} = \mathbb{E}(R^2) \text{cov} \mathbf{S} = d \text{cov} \mathbf{S} \Rightarrow \text{cov} \mathbf{S} = I_d/d. \quad (21)$$

- For  $\mathbf{Y} \sim S_d(\psi)$  with  $\mathbb{E}(R^2) < \infty$ , it follows that

$$\underset{\text{Th. 6.16}}{\text{cov}} \mathbf{Y} = \mathbb{E}(R^2) \text{cov} \mathbf{S} = \frac{\mathbb{E}(R^2)}{d} I_d$$

and thus  $\text{corr} \mathbf{Y} = I_d$ .

- For  $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Y}$  with  $\mathbb{E}(R^2) < \infty$  and Cholesky factor  $A$  of a covariance matrix  $\Sigma$ , we have  $\text{cov} \mathbf{X} = \frac{\mathbb{E}(R^2)}{d} \Sigma$  and  $\text{corr} \mathbf{X} = P$  (the correlation matrix corresponding to  $\Sigma$ ).

### Example 6.20 ( $t$ distribution)

For  $\mathbf{Y} \sim t_d(\nu, \mathbf{0}, I_d)$ ,  $R^2 = \mathbf{Y}'\mathbf{Y} = W\mathbf{Z}'\mathbf{Z}$  for  $\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$ . Therefore,

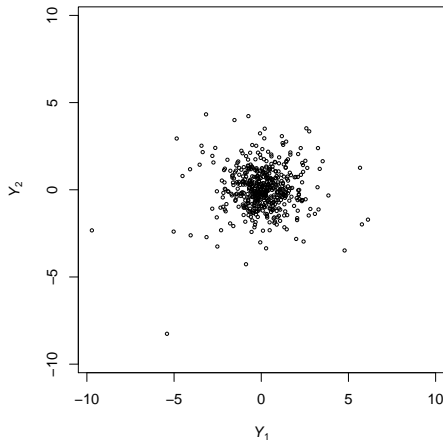
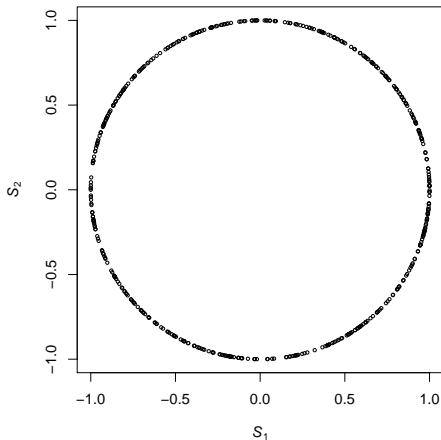
$$\frac{R^2}{d} = \frac{\mathbf{Z}'\mathbf{Z}/d}{(\nu/W)/\nu} = \frac{\chi_d^2/d}{\chi_\nu^2/\nu} \sim F(d, \nu)$$

and thus  $\mathbb{E}(R^2/d) = \frac{\nu}{\nu-2}$ .

- It follows that  $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$  has  $\text{cov } \mathbf{X} = \frac{\nu}{\nu-2}\Sigma$  and  $\text{corr } \mathbf{X} = P$  (which we already know from Section 6.2.1); note that in the univariate case  $X \sim t(\nu, \mu, \sigma^2)$ ,  $\text{var}(X) = \frac{\nu}{\nu-2}\sigma^2$ .
- We also see that we can use a Q-Q plot of the order statistics of  $R^2/d = \|\mathbf{Y}\|^2/d$  versus the theoretical quantiles of a (hypothesized)  $F(d, \nu)$  distribution to check the goodness-of-fit of the hypothesized  $t$  distribution (in any dimensions).
- See the appendix for the form of the density generator  $g$ .

### Example 6.21 (Understanding spherical distributions)

$n = 500$  realizations of  $\mathbf{S}$  (left) and  $\mathbf{Y} = R\mathbf{S}$  (right) for  $R \sim \sqrt{dF(d, \nu)}$ ,  $d = 2$ ,  $\nu = 4$  (as for the multivariate  $t_4$  distribution).



## 6.3.2 Elliptical distributions

### Definition 6.22 (Elliptical distribution)

A random vector  $\mathbf{X} = (X_1, \dots, X_d)$  has an *elliptical distribution* if

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}, \quad (\text{multivariate affine transformation})$$

where  $\mathbf{Y} \sim S_k(\psi)$ ,  $\mathbf{A} \in \mathbb{R}^{d \times k}$  (*scale matrix*  $\Sigma = \mathbf{A}\mathbf{A}'$ ), and (*location vector*)  $\boldsymbol{\mu} \in \mathbb{R}^d$ .

- By Theorem 6.16, an elliptical random vector **admits the stochastic representation**  $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \mathbf{R}\mathbf{A}\mathbf{S}$ , with  $\mathbf{R}$  and  $\mathbf{S}$  as before.
- The **cf** of an elliptical random vector  $\mathbf{X}$  is  $\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}(e^{it'\mathbf{X}}) = \mathbb{E}(e^{it'(\boldsymbol{\mu} + \mathbf{A}\mathbf{Y})}) = e^{it'\boldsymbol{\mu}} \mathbb{E}(e^{i(\mathbf{A}'\mathbf{t})'\mathbf{Y}}) = e^{it'\boldsymbol{\mu}} \psi(\mathbf{t}'\Sigma\mathbf{t})$ . Notation:  $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$  ( $= E_d(\boldsymbol{\mu}, c\Sigma, \psi(\cdot/c))$ ,  $c > 0$ ).
- If  $\Sigma$  is positive definite with Cholesky factor  $\mathbf{A}$ , then  $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$  if and only if  $\mathbf{Y} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim S_d(\psi)$ .

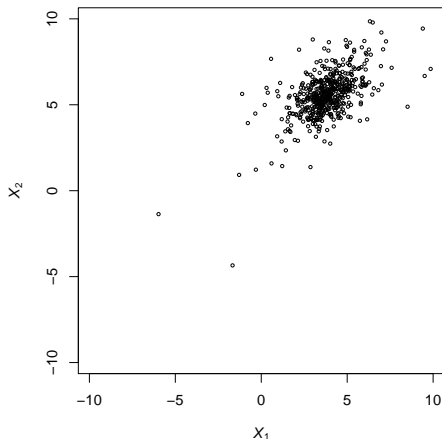
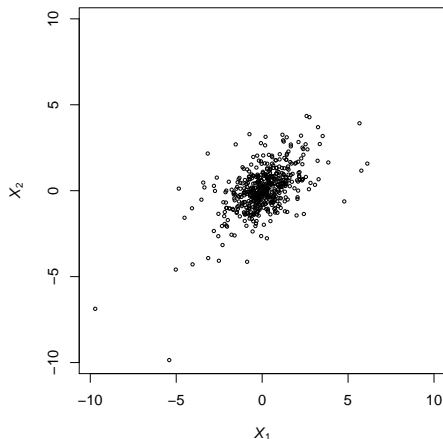
- Normal variance mixture distributions are (all) elliptical (most useful examples) since  $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + \sqrt{W} \mathbf{A} \mathbf{Z} = \boldsymbol{\mu} + \sqrt{W} \|\mathbf{Z}\| \mathbf{A} \mathbf{Z} / \|\mathbf{Z}\| = \boldsymbol{\mu} + \mathbf{R} \mathbf{A} \mathbf{S}$  with  $\mathbf{R} = \sqrt{W} \|\mathbf{Z}\|$  and  $\mathbf{S} = \mathbf{Z} / \|\mathbf{Z}\|$ . By Corollary 6.17,  $\mathbf{R}$  and  $\mathbf{S}$  are indeed independent.
- If  $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$  with  $\mathbb{P}(\mathbf{X} = \boldsymbol{\mu}) = 0$ , then  $\mathbf{Y} = \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim S_d(\psi)$ . Corollary 6.17 implies that

$$\left( \sqrt{(\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})}, \frac{\mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})}{\sqrt{(\mathbf{X} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{X} - \boldsymbol{\mu})}} \right) \stackrel{d}{=} (\mathbf{R}, \mathbf{S}), \quad (22)$$

which can be used for testing elliptical symmetry. One can also use the following result for testing.

### Example 6.23 (Understanding elliptical distributions)

$n = 500$  realizations of  $\mathbf{X} = \mathbf{RAS}$  (left) and  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{RAS}$  (right) for  $R \sim \sqrt{dF(d, \nu)}$ ,  $d = 2$ ,  $\nu = 4$ ; based on the same samples as in Example 6.21.



### 6.3.3 Properties of elliptical distributions

- **Density:** Let  $\Sigma$  be positive definite and  $\mathbf{Y} \sim S_d(\psi)$  have density generator  $g$ . The **Density Transformation Theorem** implies that  $\mathbf{X} = \boldsymbol{\mu} + \mathbf{A}\mathbf{Y}$  has density

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{\det \Sigma}} g((\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})),$$

which depends on  $\mathbf{x}$  only through  $(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$ , i.e. is constant on ellipsoids (hence the name “elliptical”).

- **Linear combinations:** For  $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$ ,  $\mathbf{B} \in \mathbb{R}^{k \times d}$  and  $\mathbf{b} \in \mathbb{R}^k$ ,

$$\mathbf{B}\mathbf{X} + \mathbf{b} \sim E_k(\mathbf{B}\boldsymbol{\mu} + \mathbf{b}, \mathbf{B}\Sigma\mathbf{B}', \psi) \quad (\text{via cfs}).$$

If  $\mathbf{a} \in \mathbb{R}^d$  (take  $\mathbf{b} = \mathbf{0}$  and  $\mathbf{B} = \mathbf{a}' \in \mathbb{R}^{1 \times d}$ ),

$$\mathbf{a}'\mathbf{X} \sim E_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a}, \psi) \quad (\text{as for } N(\boldsymbol{\mu}, \Sigma)). \quad (23)$$

From  $\mathbf{a} = \mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$  we see that all marginal distributions are of the same type.



- **Marginal dfs:** As for  $N_d(\boldsymbol{\mu}, \Sigma)$ , it immediately follows that  $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2)'\sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$  satisfies  $\mathbf{X}_1 \sim E_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$  and that  $\mathbf{X}_2 \sim E_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$ ; i.e. **margins of elliptical distributions are elliptical**.
- **Conditional distributions:** One can also show that **conditional distributions of elliptical distributions are elliptical**; see Embrechts et al. (2002). For  $N_d(\boldsymbol{\mu}, \Sigma)$  the characteristic generator remains the same.
- **Quadratic forms:** (22) implies that  $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \stackrel{d}{=} R^2$ . If  $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$ ,  $R^2 \sim \chi_d^2$ ; and if  $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$ ,  $R^2/d \sim F(d, \nu)$ .
- **Convolutions:** Let  $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, \psi)$  and  $\mathbf{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, c\Sigma, \tilde{\psi})$  be **independent**. Then  **$a\mathbf{X} + b\mathbf{Y}$  is elliptically distributed** for  $a, b \in \mathbb{R}$ ,  $c > 0$ .
- **Conditional correlations remain invariant** See Proposition A.11.

We see that **many** (but not all) **nice properties of  $N_d(\boldsymbol{\mu}, \Sigma)$  are preserved**. For estimating  $\boldsymbol{\mu}$ ,  $\Sigma$ ,  $P$ , see the appendix.

## 6.4 Dimension reduction techniques

### 6.4.1 Factor models

Explain the variability of  $\mathbf{X}$  in terms of common factors.

#### Definition 6.24 ( $p$ -factor model)

$\mathbf{X}$  follows a  *$p$ -factor model* if

$$\mathbf{X} = \mathbf{a} + B\mathbf{F} + \boldsymbol{\varepsilon}, \quad (24)$$

where

- 1)  $B \in \mathbb{R}^{d \times p}$  is a *matrix of factor loadings* and  $\mathbf{a} \in \mathbb{R}^d$ ;
- 2)  $\mathbf{F} = (F_1, \dots, F_p)$  is the random vector of (*common*) *factors* with  $p < d$  and  $\boldsymbol{\Omega} := \text{cov}(\mathbf{F})$ , (*systematic risk*);
- 3)  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)$  is the random vector of *idiosyncratic error terms* with  $\mathbb{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ ,  $\Upsilon := \text{cov}(\boldsymbol{\varepsilon})$  diag.,  $\text{cov}(\mathbf{F}, \boldsymbol{\varepsilon}) = (0)$  (*idiosync. risk*).

- **Goals:** Identify or estimate  $\mathbf{F}_t$ ,  $t \in \{1, \dots, n\}$ , then model the distribution/dynamics of the (lower-dimensional) factors (instead of  $\mathbf{X}_t$ ,  $t \in \{1, \dots, n\}$ ).
- Factor models imply that  $\Sigma := \text{cov}(\mathbf{X}) = B\Omega B' + \Upsilon$ .
- With  $B^* = B\Omega^{1/2}$  and  $\mathbf{F}^* = \Omega^{-1/2}(\mathbf{F} - \mathbb{E}(\mathbf{F}))$ , we have

$$\mathbf{X} = \boldsymbol{\mu} + B^* \mathbf{F}^* + \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\mu} = \mathbb{E}(\mathbf{X})$ . We have  $\Sigma = B^*(B^*)' + \Upsilon$ . Conversely, if  $\text{cov}(\mathbf{X}) = BB' + \Upsilon$  for some  $B \in \mathbb{R}^{d \times p}$  with  $\text{rank}(B) = p < d$  and diagonal matrix  $\Upsilon$ , then  $\mathbf{X}$  has a factor-model representation for a  $p$ -dimensional  $\mathbf{F}$  and  $d$ -dimensional  $\boldsymbol{\varepsilon}$ .

- For a one-factor/equicorrelation example, see the appendix.

## 6.4.2 Statistical estimation strategies

Consider  $\mathbf{X}_t = \mathbf{a} + B\mathbf{F}_t + \varepsilon_t$ ,  $t \in \{1, \dots, n\}$ . Three types of factor model are commonly used:

- 1) *Macroeconomic factor models*: Here we assume that  $\mathbf{F}_t$  is observable,  $t \in \{1, \dots, n\}$ . Estimation of  $B, \mathbf{a}$  is accomplished by time series regression.
- 2) *Fundamental factor models*: Here we assume that the matrix of factor loadings  $B$  is known but the factors  $\mathbf{F}_t$  are unobserved (and have to be estimated from  $\mathbf{X}_t$ ,  $t \in \{1, \dots, n\}$ , using cross-sectional regression at each  $t$ ).
- 3) *Fundamental factor models*: Here we assume that neither the factors  $\mathbf{F}_t$  nor the factor loadings  $B$  are observed (both have to be estimated from  $\mathbf{X}_t$ ,  $t \in \{1, \dots, n\}$ ). The factors can be found with principal component analysis.

## 6.4.3 Estimating macroeconomic factor models

There are *two equivalent approaches*.

### Univariate regression

- Consider the (univariate) *time series regression* model

$$X_{t,j} = a_j + \mathbf{b}_j' \mathbf{F}_t + \varepsilon_{t,j}, \quad t \in \{1, \dots, n\}.$$

- To justify the use of the *ordinary least-squares (OLS)* method to derive statistical properties of the method it is usually *assumed that*, conditional on the factors, the errors  $\varepsilon_{1,j}, \dots, \varepsilon_{n,j}$  *form a white noise process* (i.e. are identically distributed and serially uncorrelated).
- $\hat{a}_j$  estimates  $a_j$ ,  $\hat{\mathbf{b}}_j$  estimates the  $j$ th row of  $B$ .

For the multivariate case, see the appendix.

## 6.4.4 Estimating fundamental factor models

- Consider the cross-sectional regression model  $\mathbf{X}_t = B\mathbf{F}_t + \boldsymbol{\varepsilon}_t$  ( $B$  known;  $\mathbf{F}_t$  to be estimated;  $\text{cov}(\boldsymbol{\varepsilon}) = \Upsilon$ ); note that  $\mathbf{a}$  can be absorbed into  $\mathbf{F}_t$ . To obtain precision in estimating  $\mathbf{F}_t$ , we need  $d \gg p$ .
- First estimate  $\mathbf{F}_t$  via OLS by  $\hat{\mathbf{F}}_t^{\text{OLS}} = (B'B)^{-1}B'\mathbf{X}_t$ . This is the best linear unbiased estimator if the  $\boldsymbol{\varepsilon}$  is homoskedastic. However, it is possible to obtain linear unbiased estimates with a smaller covariance matrix via generalized least squares (GLS).
- To this end, estimate  $\Upsilon$  by  $\hat{\Upsilon}$  via the diagonal of the sample covariance matrix of the residuals  $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{X}_t - B\hat{\mathbf{F}}_t^{\text{OLS}}$ ,  $t \in \{1, \dots, n\}$ .
- Then estimate  $\mathbf{F}_t$  via  $\hat{\mathbf{F}}_t = (B'\hat{\Upsilon}^{-1}B)^{-1}B'\hat{\Upsilon}^{-1}\mathbf{X}_t$ .

## 6.4.5 Principal component analysis

- **Goal:** Reduce the dimensionality of highly correlated data by finding a small number of uncorrelated linear combinations which account for most of the variance in the data; this can be used for finding factors.
- **Key:** Any symmetric  $A$  admits a *spectral decomposition*

$$A = \Gamma \Lambda \Gamma',$$

where

- 1)  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d)$  is the diagonal matrix of eigenvalues of  $A$  which, w.l.o.g., are ordered so that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ ; and
  - 2)  $\Gamma$  is an orthogonal matrix whose columns are eigenvectors of  $A$  standardized to have length 1.
- Let  $\Sigma = \Gamma \Lambda \Gamma'$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d \geq 0$  (positive semidefiniteness  $\Rightarrow$  all eigenvalues  $\geq 0$ ) and  $\mathbf{Y} = \Gamma'(\mathbf{X} - \boldsymbol{\mu})$  (the so-called *principal component transform*). The  $j$ th component  $Y_j = \boldsymbol{\gamma}_j'(\mathbf{X} - \boldsymbol{\mu})$  is the  $j$ th *principal component of  $\mathbf{X}$*  (where  $\boldsymbol{\gamma}_j$  is the  $j$ th column of  $\Gamma$ ).

- We have  $\mathbb{E}\mathbf{Y} = \mathbf{0}$  and  $\text{cov}(\mathbf{Y}) = \Gamma'\Sigma\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda$ , so the principal components are uncorrelated and  $\text{var}(Y_j) = \lambda_j$ ,  $j \in \{1, \dots, d\}$ . The principal components are thus ordered by decreasing variance.
- One can show:
  - ▶ The first principal component is that standardized linear combination of  $\mathbf{X}$  which has maximal variance among all such combinations, i.e.  $\text{var}(\gamma_1'\mathbf{X}) = \max\{\text{var}(\mathbf{a}'\mathbf{X}) : \mathbf{a}'\mathbf{a} = 1\}$ .
  - ▶ For  $j \in \{2, \dots, d\}$ , the  $j$ th principal component is that standardized linear combination of  $\mathbf{X}$  which has maximal variance among all such linear combinations which are orthogonal to (and hence uncorrelated with) the first  $j - 1$ -many linear combinations.
- $\sum_{j=1}^d \text{var}(Y_j) = \sum_{j=1}^d \lambda_j = \text{trace}(\Sigma) = \sum_{j=1}^d \text{var}(X_j)$ , so we can interpret  $\sum_{j=1}^k \lambda_j / \sum_{j=1}^d \lambda_j$  as the fraction of total variance explained by the first  $k$  principal components.



## Principal components as factors

- Inverting the principal component transform  $\mathbf{Y} = \Gamma'(\mathbf{X} - \boldsymbol{\mu})$ , we have

$$\mathbf{X} = \boldsymbol{\mu} + \Gamma\mathbf{Y} = \boldsymbol{\mu} + \Gamma_1\mathbf{Y}_1 + \Gamma_2\mathbf{Y}_2 =: \boldsymbol{\mu} + \Gamma_1\mathbf{Y}_1 + \boldsymbol{\varepsilon}$$

where  $\mathbf{Y}_1 \in \mathbb{R}^k$  contains the first  $k$  principal components. This is reminiscent of the basic factor model.

- Although  $\varepsilon_1, \dots, \varepsilon_d$  will tend to have small variances, the assumptions of the factor model are generally violated (since they need not have a diagonal covariance matrix and need not be uncorrelated with  $\mathbf{Y}_1$ ). Nevertheless, principal components are often interpreted as factors.
- In principle, the same can be applied to the sample covariance matrix to obtain the sample principal components; see the appendix.