6 Multivariate models

6.1 Basics of multivariate modelling

6.1.1 Random vectors and their distributions

Joint and marginal distributions

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

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

- The *jth margin* or *marginal df F_j of $\boldsymbol{X}$* is

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

(interpreted as a limit).

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

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

- *F is absolutely continuous* if

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

for some $f \ge 0$ known as the *(joint) density of $\boldsymbol{X}$ (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 $f_j \ge 0$ known as the *density of X_j* (or F_j).

- In case f exists, $F_j(x_j) \underset{(*)}{=} \int_{-\infty}^{x_j} \int_{(-\boldsymbol{\infty}, \boldsymbol{\infty})} f(\boldsymbol{z})\, d\boldsymbol{z}_{-j}\, dz_j = \int_{-\infty}^{x_j} f_j(z_j)\, dz_j$, 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 $\boldsymbol{X}$*,

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

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

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

- Note that $\bar{F}(\boldsymbol{x}) \neq 1 - F(\boldsymbol{x})$ 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 $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2') \sim F$. The *conditional df of $\boldsymbol{X}_2$ given $\boldsymbol{X}_1 = \boldsymbol{x}_1$* is $F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1) = \mathbb{P}(\boldsymbol{X}_2 \le \boldsymbol{x}_2 \,|\, \boldsymbol{X}_1 = \boldsymbol{x}_1) = \mathbb{E}(I_{\{\boldsymbol{X}_2 \le \boldsymbol{x}_2\}} \,|\, \boldsymbol{X}_1 = \boldsymbol{x}_1)$, where $\mathbb{E}(\cdot \,|\, \cdot)$ denotes conditional expectation (not discussed here).
- A useful identity for conditional dfs is

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

  see the appendix for a proof.
  - If $\boldsymbol{x}_1 \to \infty$, then $F_{\boldsymbol{X}_2}(\boldsymbol{x}_2) = \int_{\mathbb{R}^d} F_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{z}) \, dF_{\boldsymbol{X}_1}(\boldsymbol{z})$.
  - If F has a density f, then $f_{\boldsymbol{X}_2}(\boldsymbol{x}_2) = \int_{\mathbb{R}^d} f_{\boldsymbol{X}_2|\boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{z}) \, dF_{\boldsymbol{X}_1}(\boldsymbol{z})$.

- If F has density f and $f_{\boldsymbol{X}_1}$ denotes the density of $\boldsymbol{X}_1$, then

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{\partial^2}{\partial \boldsymbol{x}_2 \partial \boldsymbol{x}_1} F(\boldsymbol{x}_1, \boldsymbol{x}_2) \underset{(14)}{=} \frac{\partial}{\partial \boldsymbol{x}_2} F_{\boldsymbol{X}_2 | \boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1) f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)$$
$$= f_{\boldsymbol{X}_2 | \boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1) f_{\boldsymbol{X}_1}(\boldsymbol{x}_1).$$

  We call

$$f_{\boldsymbol{X}_2 | \boldsymbol{X}_1}(\boldsymbol{x}_2 \,|\, \boldsymbol{x}_1) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f_{\boldsymbol{X}_1}(\boldsymbol{x}_1)}$$

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

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

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

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

Moments and characteristic function

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

$$\mathbb{E}\boldsymbol{X} = (\mathbb{E}X_1, \ldots, \mathbb{E}X_d).$$

  One can show: $X_1, \ldots, X_d$ 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 $\boldsymbol{X}$* is defined by

$$\operatorname{cov}(\boldsymbol{X}) = \mathbb{E}((\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})').$$

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

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

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

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

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

  which is in $[-1, 1]$ with $\operatorname{corr}(X_i, X_j) = \pm 1$ if and only if $X_j \stackrel{\text{a.s.}}{=} aX_i + b$ for some $a \gtrless 0$ and $b \in \mathbb{R}$.
- **Some properties of $\mathbb{E}()$ and $\operatorname{cov}()$:**

  1) For all $A \in \mathbb{R}^{k \times d}$, $\boldsymbol{b} \in \mathbb{R}^k$:
     - $\mathbb{E}(A\boldsymbol{X} + \boldsymbol{b}) = A\mathbb{E}\boldsymbol{X} + \boldsymbol{b} = A\boldsymbol{\mu} + \boldsymbol{b}$;

- $\mathrm{cov}(A\boldsymbol{X}+\boldsymbol{b}) = A\,\mathrm{cov}(\boldsymbol{X})A' = A\Sigma A'$; if $k=1$ ($A=\boldsymbol{a}'$),

$$\boldsymbol{a}'\Sigma\boldsymbol{a} = \mathrm{cov}(\boldsymbol{a}'\boldsymbol{X}) = \mathrm{var}(\boldsymbol{a}'\boldsymbol{X}) \geq 0, \quad \boldsymbol{a}\in\mathbb{R}^d, \tag{15}$$

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

- $\mathrm{cov}(\boldsymbol{X}_1+\boldsymbol{X}_2) = \mathrm{cov}(\boldsymbol{X}_1)+\mathrm{cov}(\boldsymbol{X}_2)+2\,\mathrm{cov}(\boldsymbol{X}_1,\boldsymbol{X}_2)$

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

3) A symmetric, positive (semi)definite Σ can be written as

$$\Sigma = AA' \quad \textit{Cholesky decomposition} \tag{16}$$

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 $\boldsymbol{X}$ can often be shown with the *characteristic function (cf)*

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

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

Proposition 6.1 (Characterization of covariance matrices)
A symmetric matrix Σ is a covariance matrix if and only if it is symmetric and positive semidefinite.

Proof.

"$\Rightarrow$" As we have seen in (15), a covariance matrix Σ is positive semidefinite.

"$\Leftarrow$" Let Σ be positive semidefinite with Cholesky factor A. Let $\boldsymbol{X}$ be a random vector with $\operatorname{cov}\boldsymbol{X} = I_d = \operatorname{diag}(1,\dots,1)$ (e.g. $X_j \overset{\text{ind.}}{\sim} \mathrm{N}(0,1)$). Then $\operatorname{cov}(A\boldsymbol{X}) = A\operatorname{cov}(\boldsymbol{X})A' = AA' = \Sigma$, i.e. Σ is a covariance matrix (namely that of $A\boldsymbol{X}$). □

6.1.2 Standard estimators of covariance and correlation

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

- Non-parametric method-of-moments-like estimators of $\boldsymbol{\mu}, \Sigma, P$ are

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

$$S = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})' \; (\textit{sample covariance matrix})$$

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

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

$$S_n = \frac{n}{n-1}S.$$

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

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

- S_n is unbiased since

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

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

6.1.3 The multivariate normal distribution

Definition 6.2 (Multivariate normal distribution)
$\boldsymbol{X} = (X_1, \dots, X_d)$ has a *multivariate normal* (or *Gaussian*) *distribution* if

$$\boldsymbol{X} \stackrel{\text{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Z}, \tag{17}$$

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

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

Proposition 6.3 (Cf of the multivariate normal distribution)
Let $\boldsymbol{X}$ be as in (17) and $\Sigma = AA'$. Then the cf of $\boldsymbol{X}$ is

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

Idea of proof. Using the fact that $\phi_Z(t) = \exp(-t^2/2)$ for $Z \sim \mathrm{N}(0,1)$ (see the appendix for a proof), we obtain that

$$\begin{aligned}
\phi_{\boldsymbol{X}}(\boldsymbol{t}) &= \mathbb{E}(\exp(i\boldsymbol{t}'(\boldsymbol{\mu} + A\boldsymbol{Z}))) \underset{\tilde{\boldsymbol{t}}'=\boldsymbol{t}'A}{=} \exp(i\boldsymbol{t}'\boldsymbol{\mu})\mathbb{E}(\exp(i\tilde{\boldsymbol{t}}'\boldsymbol{Z}))\\
&\overset{\text{ind.}}{=} \exp(i\boldsymbol{t}'\boldsymbol{\mu}) \prod_{j=1}^{d} \mathbb{E}(\exp(i(\tilde{t}_j Z_j))) = \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\sum_{j=1}^{d} \tilde{t}_j^2\right)\\
&= \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\tilde{\boldsymbol{t}}'\tilde{\boldsymbol{t}}\right) = \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'AA'\boldsymbol{t}\right)\\
&= \exp\left(i\boldsymbol{t}'\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right) \qquad \square
\end{aligned}$$

- We see that the multivariate normal distribution is characterized by $\boldsymbol{\mu}$ and Σ, hence the notation $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.
- $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ can be characterized by univariate normal distributions.

Proposition 6.4 (Characterization of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$)
$\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \iff \boldsymbol{a}'\boldsymbol{X} \sim \mathrm{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}) \quad \forall\, \boldsymbol{a} \in \mathbb{R}^d$.

Proof. "$\Rightarrow$" via uniqueness of cfs; "$\Leftarrow$" via Corollary A.10 $\square$

Consequences:

- Margins: $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \underset{\not\Leftarrow}{\overset{\boldsymbol{a}=\boldsymbol{e}_j}{\Rightarrow}} X_j \sim \mathrm{N}(\mu_j, \sigma_{jj}^2), \quad j \in \{1, \dots, d\}$.
- Sums: $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \overset{\boldsymbol{a}=\boldsymbol{1}}{\Rightarrow} \sum_{j=1}^d X_j \sim \mathrm{N}(\sum_{j=1}^d \mu_j,\ \sum_{i,j} \sigma_{ij})$.

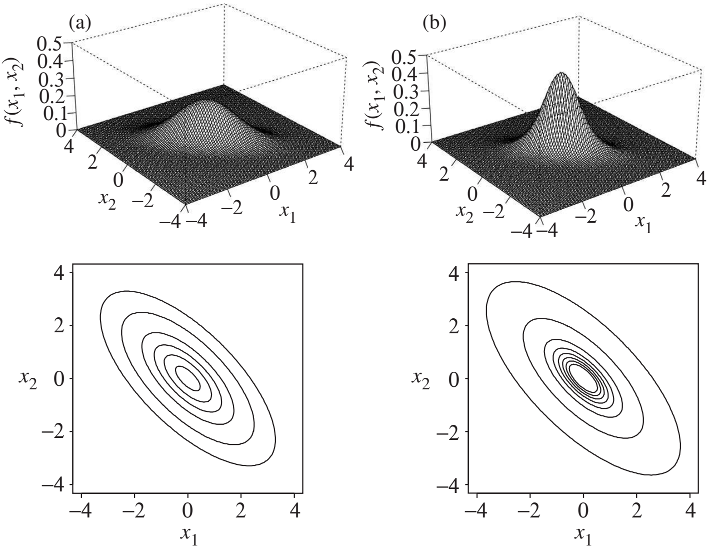
Proposition 6.5 (Density)

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

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

Consequences:

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



Left: $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ 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 $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ in terms of a stochastic representation ($\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + A\boldsymbol{Z}$) directly justifies the following sampling algorithm.

Algorithm 6.6 (Sampling $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$)

Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ symmetric and positive definite.

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

Further useful properties of multivariate normal distributions

- **Linear combinations**

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

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

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

- **Marginal dfs**

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

$$\boldsymbol{X}_1 \sim \mathrm{N}_k(\boldsymbol{\mu}_1, \Sigma_{11}) \quad \text{and} \quad \boldsymbol{X}_2 \sim \mathrm{N}_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}).$$

  Proof. Choose $B = \left(\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $B = \left(\begin{smallmatrix} 0 & 0 \\ 0 & I_{d-k} \end{smallmatrix}\right)$, respectively, in the above.

- **Conditional distributions**

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

$$\boldsymbol{X}_2 \,|\, \boldsymbol{X}_1 = \boldsymbol{x}_1 \sim \mathrm{N}_{d-k}(\boldsymbol{\mu}_{2.1}, \Sigma_{22.1}),$$

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

- **Quadratic forms**

  Let $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ and Σ be positive definite with Cholesky factor A.

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

$$(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) = \boldsymbol{Z}'\boldsymbol{Z} \sim \chi^2_d, \tag{18}$$

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

- **Convolutions**

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

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

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,

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{ind.}}{\sim} \mathrm{N}_d(\boldsymbol{\mu}, \Sigma) \;\Rightarrow\; \boldsymbol{a}'\boldsymbol{X}_1, \dots, \boldsymbol{a}'\boldsymbol{X}_n \overset{\text{ind.}}{\sim} \mathrm{N}(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}).$$

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

$\boldsymbol{a} = \boldsymbol{e}_j$, $j \in \{1, \dots, d\}$, 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 $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ with Σ positive definite, then $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2_d$.
  - Let $D_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})$ denote the *squared Mahalanobis distances* and $D_{ij} = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})'S^{-1}(\boldsymbol{X}_j - \bar{\boldsymbol{X}})$ 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 $n \to \infty$,

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

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

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

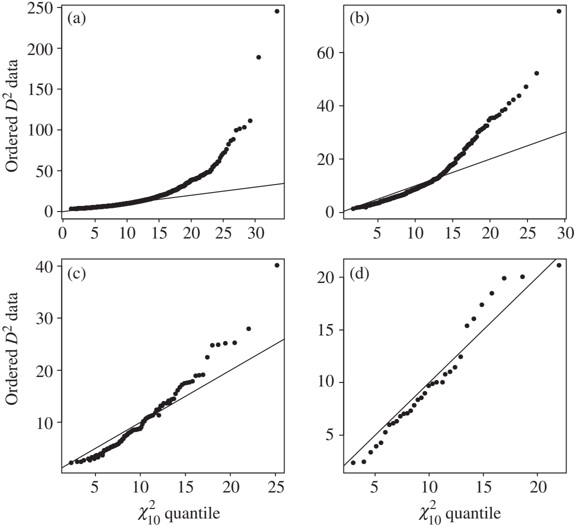
- We apply Mardia's test (of multivariate skewness and kurtosis) to daily/weekly/monthly/quarterly log-returns of 10 (of the 30) Dow Jones stocks from 1993–2000.

	Daily	Weekly	Monthly	Quarterly
n	2020	416	96	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 χ^2_{10} using a Q-Q plot; see the next page.

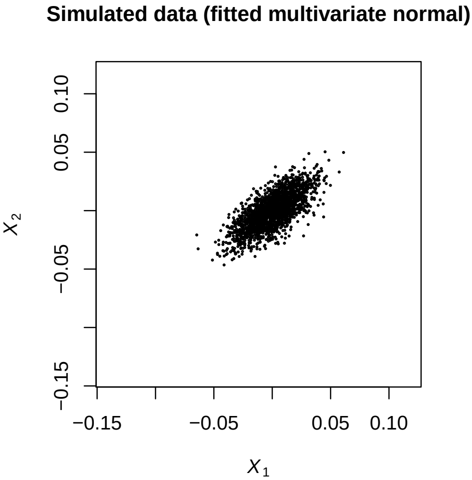
Conclusion: Daily/weekly/monthly data: Evidence against joint normality; Quarterly data: CLT effect seems to take place (but too little data to say more); still evidence against joint normality.

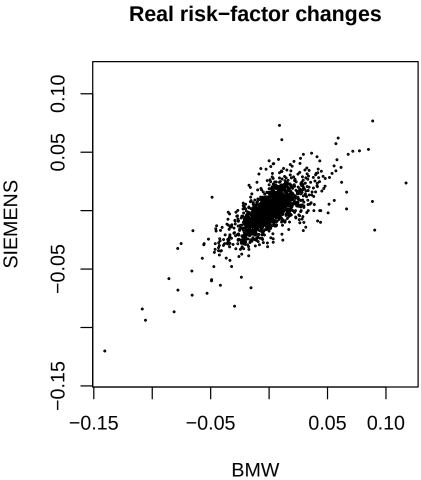
Q-Q plot of D_i^2 data against a χ^2_{10} distribution:
(a) daily data; (b) weekly data; (c) monthly data; and (d) quarterly data



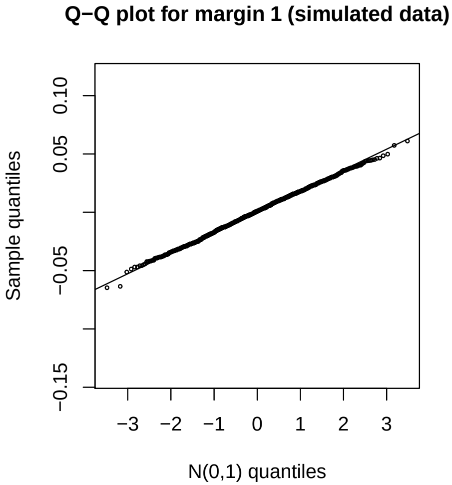
Example 6.8 (Simulated data vs BMW–Siemens)

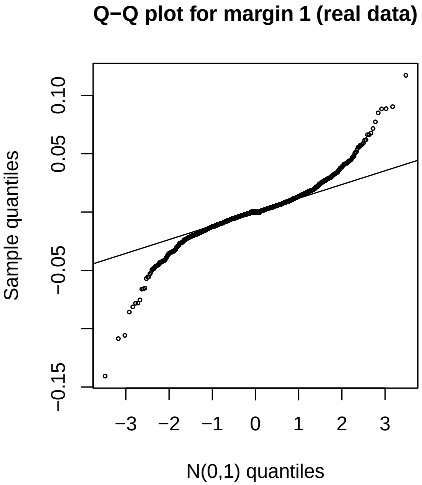
Is the BMW–Siemens data (see Section 3.2.2) jointly normal?



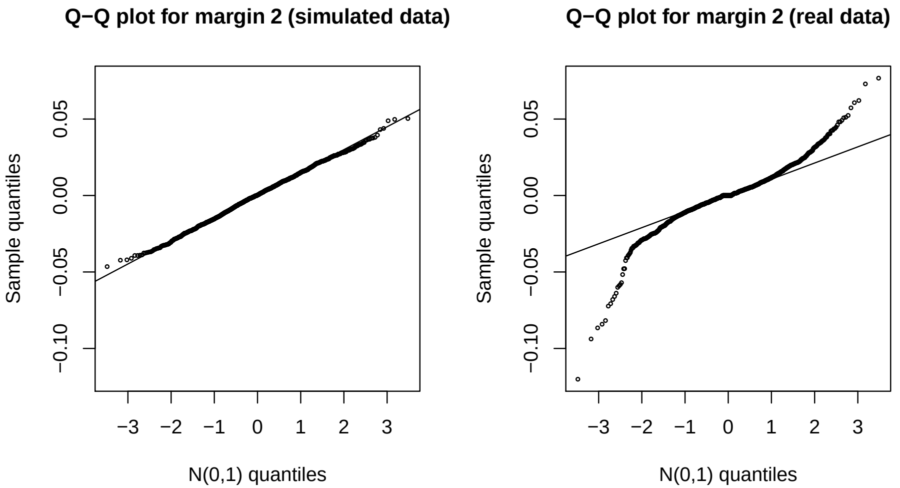


Considering the first margin only:

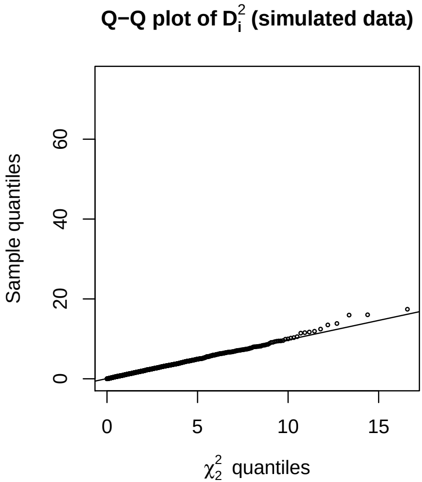


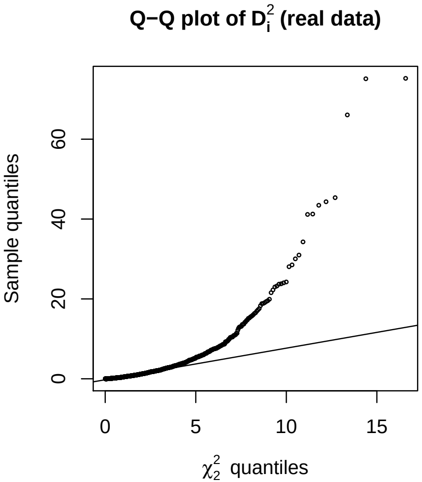


Considering the second margin only:



Q-Q plot of the simulated (left) or real (right) D_i^2's against a χ_2^2:





Advantages of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$

- Inference "easy".
- Distribution is determined by $\boldsymbol{\mu}$ and Σ.
- Linear combinations are normal ($\Rightarrow$ VaR_α and ES_α 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 $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ 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). $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ cannot capture the notion of tail dependence (see Chapter 7).

3) Very strong symmetry known as radial symmetry: $\boldsymbol{X}$ is called *radially symmetric about $\boldsymbol{\mu}$* if $\boldsymbol{X} - \boldsymbol{\mu} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} - \boldsymbol{X}$. This is true for $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$.

Short outlook:

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

6.2 Normal mixture distributions

Idea: Randomize Σ (and $\boldsymbol{\mu}$) with a non-negative rv W.

6.2.1 Normal variance mixtures

Definition 6.9 (Multivariate normal variance mixtures)
The random vector $\boldsymbol{X}$ has a (multivariate) *normal variance mixture distribution* if

$$\boldsymbol{X} \overset{\text{d}}{=} \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}, \qquad (19)$$

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

Observe that $(\boldsymbol{X} \mid W = w) \overset{\text{d}}{=} \boldsymbol{\mu} + \sqrt{w} A \boldsymbol{Z} = \mathrm{N}_d(\boldsymbol{\mu}, wAA') = \mathrm{N}_d(\boldsymbol{\mu}, w\Sigma)$; or $(\boldsymbol{X} \mid W) \overset{\text{d}}{=} \mathrm{N}_d(\boldsymbol{\mu}, W\Sigma)$. W can be interpreted as a shock affecting the variances of all risk factors.

Properties of multivariate normal variance mixtures

Let $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W} A \boldsymbol{Z}$ and $\boldsymbol{Y} = \boldsymbol{\mu} + A\boldsymbol{Z}$. Assume that $\operatorname{rank}(A) = d \le k$ and that Σ is positive definite.

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

$$
\begin{aligned}
\operatorname{cov}(\boldsymbol{X}) &= \operatorname{cov}(\sqrt{W} A \boldsymbol{Z}) = \mathbb{E}((\sqrt{W} A \boldsymbol{Z})(\sqrt{W} A \boldsymbol{Z})') \\
&\overset{\text{ind.}}{=} \mathbb{E}(W) \cdot \mathbb{E}(A \boldsymbol{Z} \boldsymbol{Z}' A') = \mathbb{E}(W) \cdot A \mathbb{E}(\boldsymbol{Z} \boldsymbol{Z}') A' \\
&= \mathbb{E}(W) A I_k A' = \mathbb{E}(W)\Sigma \underset{\text{in general}}{\ne} \Sigma \quad (= \operatorname{cov}(\boldsymbol{Y}))
\end{aligned}
$$

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

Lemma 6.10 (Independence in normal variance mixtures)
Let $\boldsymbol{X} = \boldsymbol{\mu} + \sqrt{W}\boldsymbol{Z}$ with $\mathbb{E}W < \infty$ (uncorrelated normal variance mixture). Then

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

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

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

Furthermore, $\boldsymbol{X} \,|\, W = w \sim \mathrm{N}_d(\boldsymbol{\mu}, w\Sigma)$

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

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

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

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

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

$\Rightarrow$ Only depends on $\boldsymbol{x}$ through $(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$.
$\Rightarrow$ Multivariate normal variance mixtures are elliptical distributions.

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

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

- **Sampling:**

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

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

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

For Step 2), generate $V \sim \chi^2_\nu$ and set $W = \frac{\nu}{V} \sim \mathrm{Ig}(\nu/2, \nu/2)$; 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**

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

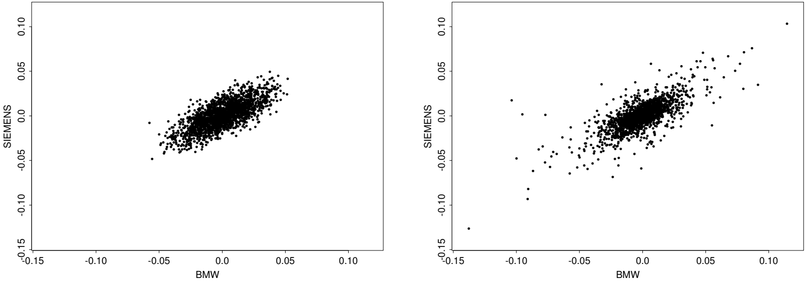
  Can be used to model ordinary and stress regimes; extends to k regimes.
- **Symmetric generalised hyperbolic distribution**
  W has a generalised inverse Gaussian distribution (GIG); see McNeil et al. (2015, p. 187)
- **Multivariate t distribution**
  W has an inverse gamma distribution $W = 1/V$ for $V \sim \Gamma(\nu/2, \nu/2)$.
  - $\mathbb{E}(W) = \frac{\nu}{\nu-2} \Rightarrow \operatorname{cov}(\boldsymbol{X}) = \frac{\nu}{\nu-2}\Sigma$. 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

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

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

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

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

$$\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{m}(W) + \sqrt{W} A \boldsymbol{Z}, \tag{20}$$

where

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

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

Example 6.13

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

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

we have

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

- If W has a GIG distribution, then $\boldsymbol{X}$ 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 $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ 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 $\boldsymbol{Y} = (Y_1, \dots, Y_d)$ 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 $UU' = U'U = I_d$)

$\boldsymbol{Y} \overset{\mathrm{d}}{=} U\boldsymbol{Y}$ (distributionally invariant under rotations and reflections)

Theorem 6.15 (Characterization of spherical distributions)

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

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

Theorem 6.16 (Stochastic representation)

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

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

Corollary 6.17

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

$$(\|\boldsymbol{Y}\|, \tfrac{\boldsymbol{Y}}{\|\boldsymbol{Y}\|}) \overset{\mathrm{d}}{=} (\|R\boldsymbol{S}\|, \tfrac{R\boldsymbol{S}}{\|R\boldsymbol{S}\|}) = (|R|\|\boldsymbol{S}\|, \tfrac{R\boldsymbol{S}}{|R|\|\boldsymbol{S}\|}) = (R, \boldsymbol{S}).$$

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

Example 6.18 (Standardized normal variance mixtures)

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

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

  so $\boldsymbol{Y} \sim S_d(\psi)$ by Theorem 6.15 Part 2). We thus have $\psi(t) = \hat{F}_W(t/2)$.
- For $\boldsymbol{Y} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)$, $\psi(t) = \exp(-t/2)$. By Corollary 6.17, simulating $\boldsymbol{S} \sim \mathrm{U}(\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| = 1\})$ can thus be done via $\boldsymbol{S} \overset{\mathrm{d}}{=} \boldsymbol{Y}/\|\boldsymbol{Y}\|$. Fang et al. (1990, pp. 48) show that ψ 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, cov, corr)

- It follows from $\boldsymbol{Y} \sim \mathrm{N}_d(\boldsymbol{0}, I_d)$ and $R^2 = \boldsymbol{Y}'\boldsymbol{Y} \sim \chi^2_d$ that

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

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

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

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

  and thus $\operatorname{corr}\boldsymbol{Y} = I_d$.

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

Example 6.20 (t distribution)

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

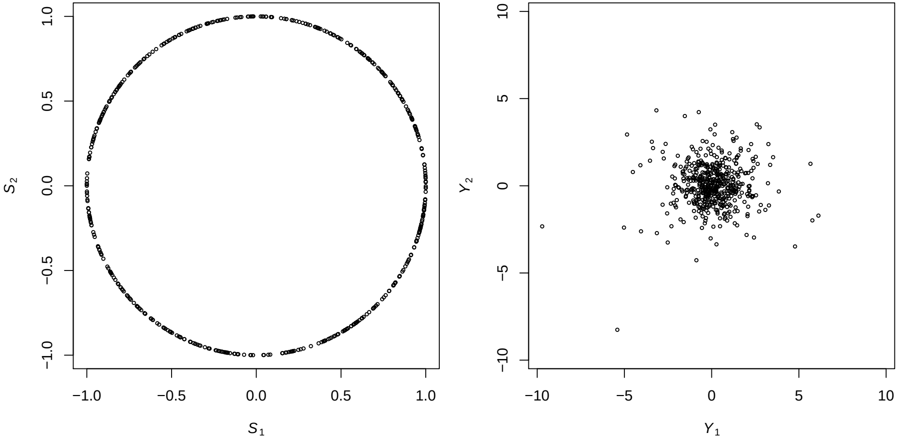
$$\frac{R^2}{d} = \frac{\boldsymbol{Z}'\boldsymbol{Z}/d}{(\nu/W)/\nu} = \frac{\chi_d^2/d}{\chi_\nu^2/\nu} \sim F(d,\nu)$$

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

- This, together with Example 6.19, implies that $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$ has $\operatorname{cov} \boldsymbol{X} = \frac{\nu}{\nu-2}\Sigma$ and $\operatorname{corr} \boldsymbol{X} = P$ (which we already know from Section 6.2.1); note that in the univariate case $X \sim t(\nu, \mu, \sigma^2)$ and $\operatorname{var}(X) = \frac{\nu}{\nu-2}\sigma^2$.
- We also see that we can use a Q-Q plot of the order statistics of $R^2/d = \|\boldsymbol{Y}\|^2/d$ 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 $\boldsymbol{S}$ (left) and $\boldsymbol{Y} = R\boldsymbol{S}$ (right) for $R \sim \sqrt{dF(d,\nu)}$, $d = 2$, $\nu = 4$ (as for the multivariate t distribution with $\nu = 4$).



6.3.2 Elliptical distributions

Definition 6.22 (Elliptical distribution)

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

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

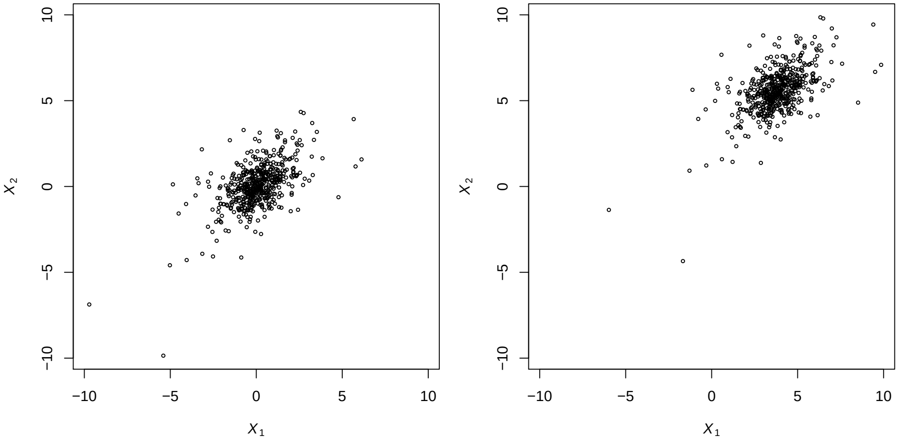
where $\boldsymbol{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{d \times k}$ (*scale matrix* $\Sigma = AA'$), and (*location vector*) $\boldsymbol{\mu} \in \mathbb{R}^d$.

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

- Normal variance mixture distributions are elliptical (most useful examples) since $\boldsymbol{X} \stackrel{\text{d}}{=} \boldsymbol{\mu} + \sqrt{W} A\boldsymbol{Z} = \boldsymbol{\mu} + \sqrt{W}\|\boldsymbol{Z}\| A\boldsymbol{Z}/\|\boldsymbol{Z}\| = \boldsymbol{\mu} + RA\boldsymbol{S}$ with $R = \sqrt{W}\|\boldsymbol{Z}\|$ and $\boldsymbol{S} = \boldsymbol{Z}/\|\boldsymbol{Z}\|$. By Corollary 6.17, R and $\boldsymbol{S}$ are indeed independent.
- If $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ with $\mathbb{P}(\boldsymbol{X} = \boldsymbol{\mu}) = 0$, then $\boldsymbol{Y} = A^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim S_d(\psi)$. Corollary 6.17 implies that
$$\left(\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})},\ \frac{A^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}{\sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}}\right) \stackrel{\text{d}}{=} (R, \boldsymbol{S}), \qquad (22)$$
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 $\boldsymbol{X} = R\boldsymbol{A}\boldsymbol{S}$ (left) and $\boldsymbol{X} = \boldsymbol{\mu} + R\boldsymbol{A}\boldsymbol{S}$ (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 Σ be positive definite and $\boldsymbol{Y} \sim S_d(\psi)$ have density generator g. The Density Transformation Theorem implies that $\boldsymbol{X} = \boldsymbol{\mu} + A\boldsymbol{Y}$ has density

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

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

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

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

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

$$\boldsymbol{a}'\boldsymbol{X} \sim \mathrm{E}_1(\boldsymbol{a}'\boldsymbol{\mu}, \boldsymbol{a}'\Sigma\boldsymbol{a}, \psi) \quad \text{(as for } \mathrm{N}(\boldsymbol{\mu}, \Sigma)\text{)}. \tag{23}$$

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

- **Marginal dfs:** As for $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, it immediately follows that $\boldsymbol{X} = (\boldsymbol{X}_1', \boldsymbol{X}_2')' \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ satisfies $\boldsymbol{X}_1 \sim \mathrm{E}_k(\boldsymbol{\mu}_1, \Sigma_{11}, \psi)$ and that $\boldsymbol{X}_2 \sim \mathrm{E}_{d-k}(\boldsymbol{\mu}_2, \Sigma_{22}, \psi)$; 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 $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ the characteristic generator remains the same.
- **Quadratic forms:** (22) implies that $(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \stackrel{\mathrm{d}}{=} R^2$. If $\boldsymbol{X} \sim \mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$, $R^2 \sim \chi^2_d$; and if $\boldsymbol{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$, $R^2/d \sim F(d, \nu)$.
- **Convolutions:** Let $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu}, \Sigma, \psi)$ and $\boldsymbol{Y} \sim \mathrm{E}_d(\tilde{\boldsymbol{\mu}}, c\Sigma, \tilde{\psi})$ be independent. Then $a\boldsymbol{X} + b\boldsymbol{Y}$ is elliptically distributed for $a, b \in \mathbb{R}$, $c > 0$.
- **Conditional correlations remain invariant** See Proposition A.11.

Many (but not all) nice properties of $\mathrm{N}_d(\boldsymbol{\mu}, \Sigma)$ are preserved. For estimating $\boldsymbol{\mu}$, Σ, P, see the appendix. The following result shows why elliptical distributions are known as the “Garden of Eden” of QRM.

Proposition 6.24 (Subadditivity of VaR in elliptical models)
Let $L_i = \boldsymbol{\lambda}_i'\boldsymbol{X}$, $\boldsymbol{\lambda}_i \in \mathbb{R}^d$, $i \in \{1,\dots,n\}$, with $\boldsymbol{X} \sim \mathrm{E}_d(\boldsymbol{\mu},\Sigma,\psi)$. Then $\mathrm{VaR}_\alpha(\sum_{i=1}^n L_i) \le \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$ for all $\alpha \in [1/2,1]$.

Proof. Consider a generic $L = \boldsymbol{\lambda}'\boldsymbol{X} \overset{\mathrm{d}}{=} \boldsymbol{\lambda}'\boldsymbol{\mu} + \boldsymbol{\lambda}'A\boldsymbol{Y}$ for $\boldsymbol{Y} \sim S_k(\psi)$. By Theorem 6.15 Part 3), $\boldsymbol{\lambda}'A\boldsymbol{Y} \overset{\mathrm{d}}{=} \|\boldsymbol{\lambda}'A\|Y_1$, so $L \overset{\mathrm{d}}{=} \boldsymbol{\lambda}'\boldsymbol{\mu} + \|\boldsymbol{\lambda}'A\|Y_1$ (all L_i's are of the same type). By translation invariance and positive homogeneity,

$$\mathrm{VaR}_\alpha(L) = \boldsymbol{\lambda}'\boldsymbol{\mu} + \|\boldsymbol{\lambda}'A\|\,\mathrm{VaR}_\alpha(Y_1). \tag{24}$$

Applying (24) once to $L = \sum_{i=1}^n L_i = (\sum_{i=1}^n \boldsymbol{\lambda}_i)'\boldsymbol{X}$ and to each $L = L_i = \boldsymbol{\lambda}_i'\boldsymbol{X}$, $i \in \{1,\dots,n\}$, and using that $\mathrm{VaR}_\alpha(Y_1) \ge 0$ for $\alpha \in [1/2,1]$, we obtain $\mathrm{VaR}_\alpha(\sum_{i=1}^n L_i) \underset{(24)}{=} \sum_{i=1}^n \boldsymbol{\lambda}_i'\boldsymbol{\mu} + \|\sum_{i=1}^n \boldsymbol{\lambda}_i'A\|\,\mathrm{VaR}_\alpha(Y_1)$
$\le \sum_{i=1}^n \boldsymbol{\lambda}_i'\boldsymbol{\mu} + (\sum_{i=1}^n \|\boldsymbol{\lambda}_i'A\|)\,\mathrm{VaR}_\alpha(Y_1) = \sum_{i=1}^n (\boldsymbol{\lambda}_i'\boldsymbol{\mu} + \|\boldsymbol{\lambda}_i'A\|\,\mathrm{VaR}_\alpha(Y_1))$
$\underset{(24)}{=} \sum_{i=1}^n \mathrm{VaR}_\alpha(L_i)$. For $\boldsymbol{\lambda}_i = \boldsymbol{e}_i$, $\mathrm{VaR}_\alpha(\sum_{i=1}^n X_i) \le \sum_{i=1}^n \mathrm{VaR}_\alpha(X_i)$. $\square$

6.4 Dimension reduction techniques

6.4.1 Factor models

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

Definition 6.25 (p-factor model)

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

$$\boldsymbol{X} = \boldsymbol{a} + B\boldsymbol{F} + \boldsymbol{\varepsilon}, \tag{25}$$

where

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

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

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

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

6.4.2 Statistical estimation strategies

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

1) *Macroeconomic factor models:* Here we assume that $\boldsymbol{F}_t$ is observable, $t \in \{1, \dots, n\}$. Estimation of $B, \boldsymbol{a}$ 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 $\boldsymbol{F}_t$ are unobserved (and have to be estimated from $\boldsymbol{X}_t$, $t \in \{1, \dots, n\}$, using cross-sectional regression at each t).

3) *Fundamental factor models:* Here we assume that neither the factors $\boldsymbol{F}_t$ nor the factor loadings B are observed (both have to be estimated from $\boldsymbol{X}_t$, $t \in \{1, \dots, n\}$). 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

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

- 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, $\hat{\boldsymbol{b}}_j$ estimates the jth row of B.

For the multivariate case, see the appendix.

6.4.4 Estimating fundamental factor models

- Consider the cross-sectional regression model $\boldsymbol{X}_t = B\boldsymbol{F}_t + \boldsymbol{\varepsilon}_t$ (B known; $\boldsymbol{F}_t$ to be estimated; $\mathrm{cov}(\boldsymbol{\varepsilon}) = \Upsilon$); note that $\boldsymbol{a}$ can be absorbed into $\boldsymbol{F}_t$. To obtain precision in estimating $\boldsymbol{F}_t$, we need $d \gg p$.

- First estimate $\boldsymbol{F}_t$ via OLS by $\hat{\boldsymbol{F}}_t^{\mathrm{OLS}} = (B'B)^{-1}B'\boldsymbol{X}_t$. This is the best linear unbiased estimator if the ε 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 Υ by $\hat{\Upsilon}$ via the diagonal of the sample covariance matrix of the residuals $\hat{\boldsymbol{\varepsilon}}_t = \boldsymbol{X}_t - B\hat{\boldsymbol{F}}_t^{\mathrm{OLS}}$, $t \in \{1, \dots, n\}$.

- Then estimate $\boldsymbol{F}_t$ via $\hat{\boldsymbol{F}}_t = (B'\Upsilon^{-1}B)^{-1}B'\Upsilon^{-1}\boldsymbol{X}_t$.

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) $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_d)$ is the diagonal matrix of eigenvalues of A which, w.l.o.g., are ordered so that $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_d$; and
2) Γ is an orthogonal matrix whose columns are eigenvectors of A standardized to have length 1.

- Let $\Sigma = \Gamma\Lambda\Gamma'$ with $\lambda_1 \ge \lambda_2 \ge \dots \ge \lambda_d \ge 0$ (positive semidefiniteness $\Rightarrow$ all eigenvalues ≥ 0) and $\boldsymbol{Y} = \Gamma'(\boldsymbol{X} - \boldsymbol{\mu})$ (the so-called *principal component transform*). The jth component $Y_j = \boldsymbol{\gamma}_j'(\boldsymbol{X} - \boldsymbol{\mu})$ is the *jth principal component of $\boldsymbol{X}$* (where $\boldsymbol{\gamma}_j$ is the jth column of Γ).

- We have $\mathbb{E}\boldsymbol{Y} = \mathbf{0}$ and $\operatorname{cov}(\boldsymbol{Y}) = \Gamma'\Sigma\Gamma = \Gamma'\Gamma\Lambda\Gamma'\Gamma = \Lambda$, so the principal components are uncorrelated and $\operatorname{var}(Y_j) = \lambda_j$, $j \in \{1, \dots, d\}$. The principal components are thus ordered by decreasing variance.
- One can show:
  - The first principal component is that standardized linear combination of $\boldsymbol{X}$ which has maximal variance among all such combinations, i.e. $\operatorname{var}(\boldsymbol{\gamma}_1'\boldsymbol{X}) = \max\{\operatorname{var}(\boldsymbol{a}'\boldsymbol{X}) : \boldsymbol{a}'\boldsymbol{a} = 1\}$.
  - For $j \in \{2, \dots, d\}$, the jth principal component is that standardized linear combination of $\boldsymbol{X}$ which has maximal variance among all such linear combinations which are orthogonal to (and hence uncorrelated with) the first $j-1$-many linear combinations.
- $\sum_{j=1}^d \operatorname{var}(Y_j) = \sum_{j=1}^d \lambda_j = \operatorname{trace}(\Sigma) = \sum_{j=1}^d \operatorname{var}(X_j)$, 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 $\boldsymbol{Y} = \Gamma'(\boldsymbol{X} - \boldsymbol{\mu})$, we have

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

  where $\boldsymbol{Y}_1 \in \mathbb{R}^k$ 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 $\boldsymbol{Y}_1$). 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.