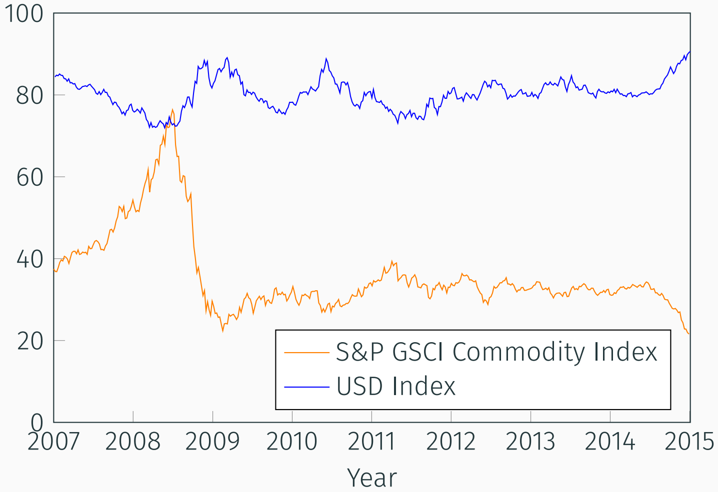
# Basics of Multivariate Modelling

Lars Christmann

June 1, 2018

Seminar on Quantitative Risk Management

# Motivation: Correlation between Commodity Prices and USD




→ Dependence structure should be incorporated into financial risk models.

Data source: Yahoo Finance

# Table of contents

# Basic Definitions and Standard Estimators

# Random Vectors and Their Distributions

$\mathbf{X} = (X_1, \dots, X_d)^T$ random vector of risk-factor changes.

- Joint distribution function
  $F(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x}) = F_{\mathbf{X}}(x_1, \dots, x_d) = P(\mathbf{X} \leqslant \mathbf{x}) = P(X_1 \leqslant x_1, \dots, X_d \leqslant x_d)$
- Marginal distribution function of $X_i$
  $F_i(x_i) = P(X_i \leqslant x_i) = F(\infty, \dots, \infty, x_i, \infty, \dots, \infty)$
- $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ where $\mathbf{X_1} = (X_1, \dots, X_k)^T$, $\mathbf{X_2} = (X_{k+1}, \dots, X_d)^T$, marginal distribution of $\mathbf{X}_1$ is
  $F_{\mathbf{X}_1}(\mathbf{x}_1) = P(\mathbf{X}_1 \leqslant \mathbf{x}_1) = F(x_1, \dots, x_k, \infty, \dots, \infty)$
- Distribution function of $\mathbf{X}$ is absolutely continuous if

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f(u_1, \dots, u_d) \, \mathrm{d}u_1 \dots \mathrm{d}u_d$$

- Existence of joint density $f$ implies existence of all $k$-dimensional marginal densities

# Conditional Distributions and Independence 1

$\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ like before.

- Density of $\mathbf{X}_2$ given $\mathbf{X}_1 = \mathbf{x_1}$

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

- Corresponding distribution function

$$F_{\mathbf{X}_2|\mathbf{X}_1}(\mathbf{x}_2|\mathbf{x}_1) = \int_{-\infty}^{x_{k+1}} \cdots \int_{-\infty}^{x_d} \frac{f(x_1, \ldots, x_k, u_{k+1}, \ldots, u_d)}{f_{\mathbf{X}_1}(\mathbf{x}_1)} \, \mathrm{d}u_{k+1} \ldots \mathrm{d}u_d$$

- If $f(\mathbf{x}) = f_{\mathbf{X}_1}(\mathbf{x}_1) f_{\mathbf{X}_2}(\mathbf{x}_2)$, then $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent

# Conditional Distributions and Independence 2

- Components of $\mathbf{X}$ are mutually independent if and only if $F(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i)$ for all $\mathbf{x} \in \mathbb{R}^d$
- When $\mathbf{X}$ possesses a density $f$, the components are mutually independent if and only if $f(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i)$ for all $\mathbf{x} \in \mathbb{R}^d$

## Mean Vectors and Covariance Matrices

- Mean vector of $\mathbf{X}$ is defined as $E(\mathbf{X}) = (E(X_1), \ldots, E(X_d))^T$
- Covariance matrix of $\mathbf{X}$ is defined as $\text{cov}(\mathbf{X}) = E\left((\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^T\right)$, where the expectation operator is applied componentwise
- Entries of the covariance matrix $\sigma_{ij} = \text{cov}(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$
- Correlation matrix $\rho(\mathbf{X})$ is defined componentwise by

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

# Properties of Mean Vectors and Covariance Matrices

Let $B \in \mathbb{R}^{k \times d}$, $\mathbf{b} \in \mathbb{R}^k$ and $\mathbf{a} \in \mathbb{R}^d$.

- $E(B\mathbf{X} + \mathbf{b}) = B\,E(\mathbf{X}) + \mathbf{b}$
- $\operatorname{cov}(B\mathbf{X} + \mathbf{b}) = B \operatorname{cov}(\mathbf{X})\, B^T$

Let $\Sigma = \operatorname{cov}(\mathbf{X})$.

- $\operatorname{var}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T \Sigma \mathbf{a} \geqslant 0$
- If $\Sigma$ is positive-definite $\rightarrow$ Cholesky decomposition $\Sigma = AA^T$
- $\Sigma^{1/2} = A$ denotes the Cholesky factor, $\Sigma^{-1/2}$ its inverse

# Standard Estimators of Covariance and Correlation 1

- $\mathbf{X}_1, \ldots, \mathbf{X}_n$ identically distributed observations of a $d$-dimensional risk-factor change
- First assumption: Independent or serially uncorrelated
- Second assumption: Their distribution has mean vector $\boldsymbol{\mu}$, finite covariance matrix $\Sigma$ and correlation matrix $P$

# Standard Estimators of Covariance and Correlation 2

- Estimator for $\mu$
  $$\bar{\mathsf{X}} := \frac{1}{n}\sum_{i=1}^{n}\mathsf{X}_i$$
  (sample mean)
- Estimator for $\Sigma$
  $$S := \frac{1}{n}\sum_{i=1}^{n}(\mathsf{X}_i - \bar{\mathsf{X}})(\mathsf{X}_i - \bar{\mathsf{X}})^T$$
  (sample covariance matrix)
- Estimator for P
  $R$ (sample correlation matrix) defined componentwise by
  $$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}s_{kk}}}$$
  where $s_{jk}$ denotes the $(j,k)$-th element of $S$

# Properties of the Standard Estimators 1

- Estimator for $\Sigma$

$$S := \frac{1}{n}\sum_{i=1}^{n}(\mathsf{X}_i - \bar{\mathsf{X}})(\mathsf{X}_i - \bar{\mathsf{X}})^T$$

  (sample covariance matrix)
- We have

$$\begin{aligned} nE(S) &= E\left(\sum_{i=1}^{n}(\mathsf{X}_i - \mu)(\mathsf{X}_i - \mu)^T - n(\bar{\mathsf{X}} - \mu)(\bar{\mathsf{X}} - \mu)^T\right) \\ &= \sum_{i=1}^{n}\operatorname{cov}(\mathsf{X}_i) - n\operatorname{cov}(\bar{\mathsf{X}}) = n\Sigma - \Sigma \end{aligned}$$

  since $\operatorname{cov}(\bar{\mathsf{X}}) = n^{-1}\Sigma$ for iid or identically distributed and uncorrelated data
- Unbiased version given by $S_u = nS/(n-1)$

# Properties of the Standard Estimators 2

- Further properties of $\bar{X}, S$ and $R$ depend on the true distribution of the observations
- If the data is iid multivariate normal, they are the maximum likelihood estimators and their behavior is well understood
- For other distributions it is less well understood and other estimators might perform better
- Not useful if the true distribution does not have a finite covariance matrix or the mean vector does not exist

# The Multivariate Normal Distribution

# Definition

$\mathsf{X} = (X_1, \ldots, X_d)^T$ has a multivariate normal distribution if

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

where $\mathsf{Z} = (Z_1, \ldots, Z_k)^T$ is a vector of iid standard normal random variables, $\mu \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times k}$. We also write $\mathsf{X} \sim N_d(\mu, \Sigma)$.

- $E(\mathsf{X}) = \mu$
- $\mathrm{cov}(\mathsf{X}) = \Sigma = AA^T$

We assume that $\Sigma$ is non-singular.
In this case $\mathsf{X}$ has an absolutely continuous distribution function with joint density

$$f(\mathsf{x}) = \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathsf{x} - \mu)^T \Sigma^{-1} (\mathsf{x} - \mu)\right\}.$$

# Joint Density of the Multivariate Normal Distribution

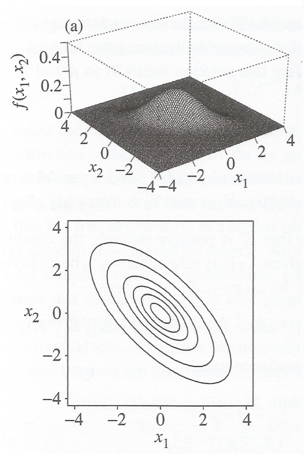



Figure 1: Density of a bivariate normal distribution with standard normal margins and correlation −0.7.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\det(\Sigma)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

- Components of $\mathbf{X}$ are mutually independent if and only if $\Sigma$ is diagonal
- Points with equal density lie on ellipsoids determined by equations of the form $(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu) = c$ for $c > 0$
- Elliptical symmetry

# Linear Combinations

Let $\mathbf{X} \sim N_d(\mu, \Sigma), B \in \mathbb{R}^{k \times d}, \mathbf{b} \in \mathbb{R}^k$ and $\mathbf{a} \in \mathbb{R}^d$.

- $B\mathbf{X} + \mathbf{b} \sim N_k(B\mu + \mathbf{b}, B\Sigma B^T)$
- In particular, $\mathbf{a}^T\mathbf{X} \sim N(\mathbf{a}^T\mu, \mathbf{a}^T\Sigma\mathbf{a})$

Alternative characterization of multivariate normality:
$\mathbf{X}$ is multivariate normal if and only if $\mathbf{a}^T\mathbf{X}$ is univariate normal for all $\mathbf{a} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$.

# Marginal Distributions and Conditional Distributions

$\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ like before. This notation can be extended, i.e.

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

- The marginal distributions of $\mathbf{X}_1$ and $\mathbf{X}_2$ are also multivariate normal, $\mathbf{X}_1 \sim N_k(\mu_1, \Sigma_{11})$ and $\mathbf{X}_2 \sim N_{d-k}(\mu_2, \Sigma_{22})$
- The conditional distribution of $\mathbf{X}_2$ given $\mathbf{X}_1 = \mathbf{x}_1$ is multivariate normal,

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \sim N_{d-k}(\mu_{2.1}, \Sigma_{22.1})$$

  where $\mu_{2.1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \mu_1)$ and $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.

# Convolutions

$X \sim N_d(\mu, \Sigma), Y \sim N_d(\mu', \Sigma')$ independent implies

$$X + Y \sim N_d(\mu + \mu', \Sigma + \Sigma').$$

# Normal margins do not imply joint normality

$X_1 \sim N(0,1)$, $Y$ independent of $X_1$ with $P(Y=1) = P(Y=-1) = \frac{1}{2}$ and $X_2 = YX_1$.

- Marginal distribution of $X_2$

$$\begin{aligned} P(X_2 \leqslant x_2) &= E\left(P(X_2 \leqslant x_2 | Y)\right) \\ &= P(Y=1)P(X_1 \leqslant x_2) + P(Y=-1)P(-X_1 \leqslant x_2) \\ &= \frac{1}{2}\Phi(x) + \frac{1}{2}\Phi(x) \end{aligned}$$

- If $(X_1, X_2)^T$ was multivariate normal, $aX_1 + bX_2$ should be univariate normal for all $a, b \in \mathbb{R} \setminus \{0\}$
- $P(X_1 + X_2 = 0) = P(Y = -1) = \frac{1}{2}$

$\Rightarrow (X_1, X_2)^T$ is not multivariate normal

# Testing Multivariate Normality

# Univariate Tests

$\mathsf{X}_1, \ldots, \mathsf{X}_n$ iid multivariate normal, then $\mathsf{a}^T\mathsf{X}_1, \ldots, \mathsf{a}^T\mathsf{X}_n$ are iid univariate normal.
$\rightarrow$ Use univariate normality tests or Q-Q plots

# Multivariate Normality Tests 1

$X \sim N_d(\mu, \Sigma)$, $Z := \Sigma^{-1/2}(X - \mu) \sim N_d(0, I_d)$

$$\Rightarrow (X - \mu)^T \Sigma^{-1} (X - \mu) = Z^T Z \sim \chi^2_d.$$

Replace $\mu$ and $\Sigma$ with our standard estimators.

$$D_i^2 = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}),\ i = 1, \dots, n$$

Under the null hypothesis $D_1^2, \dots, D_n^2$ should behave roughly like an iid $\chi^2_d$-sample. $\rightarrow$ Use Q-Q plots

# Multivariate Normality Tests 2

Mardia's tests of multinormality (based on skewness and kurtosis statistics):

$$b_d = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} D_{ij}^3, \quad k_d = \frac{1}{n}\sum_{i=1}^{n} D_i^4,$$

where $D_{ij} = (\mathsf{X}_i - \bar{\mathsf{X}})^T S^{-1}(\mathsf{X}_j - \bar{\mathsf{X}})$ and $D_i^2 = D_{ii}$.
Under the null hypothesis of multivariate normality

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

as $n \to \infty$.

# Normality of Returns on Dow Jones Stocks

- Data of ten stocks from the Dow Jones index from 1993-2000
- Daily, weekly, monthly and quarterly logarithmic returns

# Normality of Returns on Dow Jones Stocks: Mardia's Tests

| | 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 |

- Daily, weekly and monthly return data fail the tests
- Inconclusive for quarterly return data

# Normality of Returns on Dow Jones Stocks: Q-Q Plots

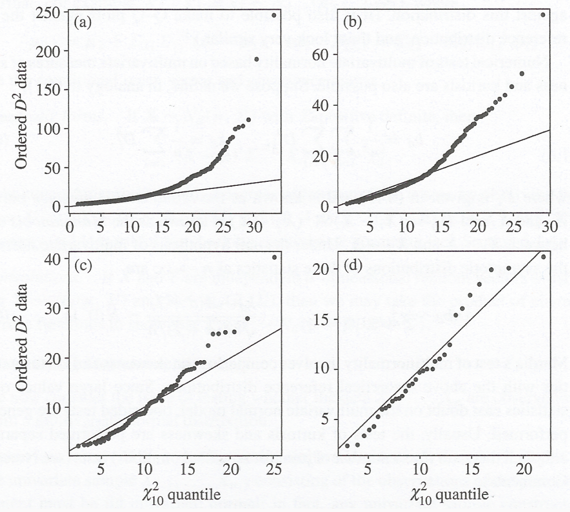



Figure 2: Q-Q plot of the $D_i^2$ data against a $\chi^2_{10}$ distribution. (a) daily analysis, (b) weekly analysis, (c) monthly analysis, (d) quarterly analysis

# Conclusion

- The multivariate normal distribution has many desirable properties that make it convenient to use
- Tests on financial return data suggest that it is not a good model in many risk-management applications
- The marginal tails are too thin
- The joint tails do not assign enough weight to joint extreme events
- The distribution has a strong form of symmetry and thus a very specific dependence structure

# References

Alexander J. McNeil, Rüdiger Frey, and Paul Embrechts.
*Quantitative Risk Management: Concepts, Techniques and Tools.*
Princeton University Press, Princeton, NJ, USA, 2015.