QUANTITATIVE RISK MANAGEMENT.

CONCEPTS, TECHNIQUES AND TOOLS

Paul Embrechts

ETH Zürich

www.math.ethz.ch/~embrechts

Contents

A. Risk Management Basics

1. Risks, Losses and Risk Factors

2. Example: Portfolio of Stocks

3. Conditional and Unconditional Loss Distributions

4. Risk Measures

5. Linearisation of Loss

6. Example: European Call Option

A1. Risks, Losses and Risk Factors

We concentrate on the following sources of risk.

- **Market Risk** - risk associated with fluctuations in value of traded assets.

- **Credit Risk** - risk associated with uncertainty that debtors will honour their financial obligations

- **Operational Risk** - risk associated with possibility of human error, IT failure, dishonesty, natural disaster etc.

This is a non-exhaustive list; other sources of risk such as liquidity risk possible.

Modelling Financial Risks

To model risk we use language of probability theory. Risks are represented by random variables mapping unforeseen future states of the world into values representing profits and losses.

The risks which interest us are aggregate risks. In general we consider a portfolio which might be

- a collection of stocks and bonds;
- a book of derivatives;
- a collection of risky loans;
- a financial institution's overall position in risky assets.

Portfolio Values and Losses

Consider a portfolio and let V_t denote its value at time t; we assume this random variable is observable at time t.

Suppose we look at risk from perspective of time t and we consider the time period $[t, t+1]$. The value V_{t+1} at the end of the time period is unknown to us.

The distribution of $(V_{t+1} - V_t)$ is known as the profit-and-loss or P&L distribution. We denote the loss by $L_{t+1} = -(V_{t+1} - V_t)$. By this convention, losses will be positive numbers and profits negative.

We refer to the distribution of L_{t+1} as the loss distribution.

Introducing Risk Factors

The Value V_t of the portfolio/position will be modelled as a function of time and a set of d underlying risk factors. We write

$$V_t = f(t, \mathbf{Z}_t) \tag{1}$$

where $\mathbf{Z}_t = (Z_{t,1}, \ldots, Z_{t,d})'$ is the risk factor vector. This representation of portfolio value is known as a mapping. Examples of typical risk factors:

- (logarithmic) prices of financial assets
- yields
- (logarithmic) exchange rates

Risk Factor Changes

We define the time series of risk factor changes by

$$\mathbf{X}_t := \mathbf{Z}_t - \mathbf{Z}_{t-1}.$$

Typically, historical risk factor time series are available and it is of interest to relate the changes in these underlying risk factors to the changes in portfolio value.

We have

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -\left(f(t+1, \mathbf{Z}_{t+1}) - f(t, \mathbf{Z}_t)\right) \\ &= -(f(t+1, \mathbf{Z}_t + \mathbf{X}_{t+1}) - f(t, \mathbf{Z}_t)) \qquad (2) \end{aligned}$$

The Loss Operator

Since the risk factor values $\mathbf{Z}_t$ are known at time t the loss L_{t+1} is determined by the risk factor changes $\mathbf{X}_{t+1}$.

Given realisation $\mathbf{z}_t$ of $\mathbf{Z}_t$, the loss operator at time t is defined as

$$l_{[t]}(\mathbf{x}) := -(f(t+1, \mathbf{z}_t + \mathbf{x}) - f(t, \mathbf{z}_t)), \tag{3}$$

so that

$$L_{t+1} = l_{[t]}(\mathbf{X}_{t+1}).$$

From the perspective of time t the loss distribution of L_{t+1} is determined by the multivariate distribution of $\mathbf{X}_{t+1}$.

But which distribution exactly? Conditional distribution of L_{t+1} given history up to and including time t or unconditional distribution under assumption that $(\mathbf{X}_t)$ form stationary time series?

A2. Example: Portfolio of Stocks

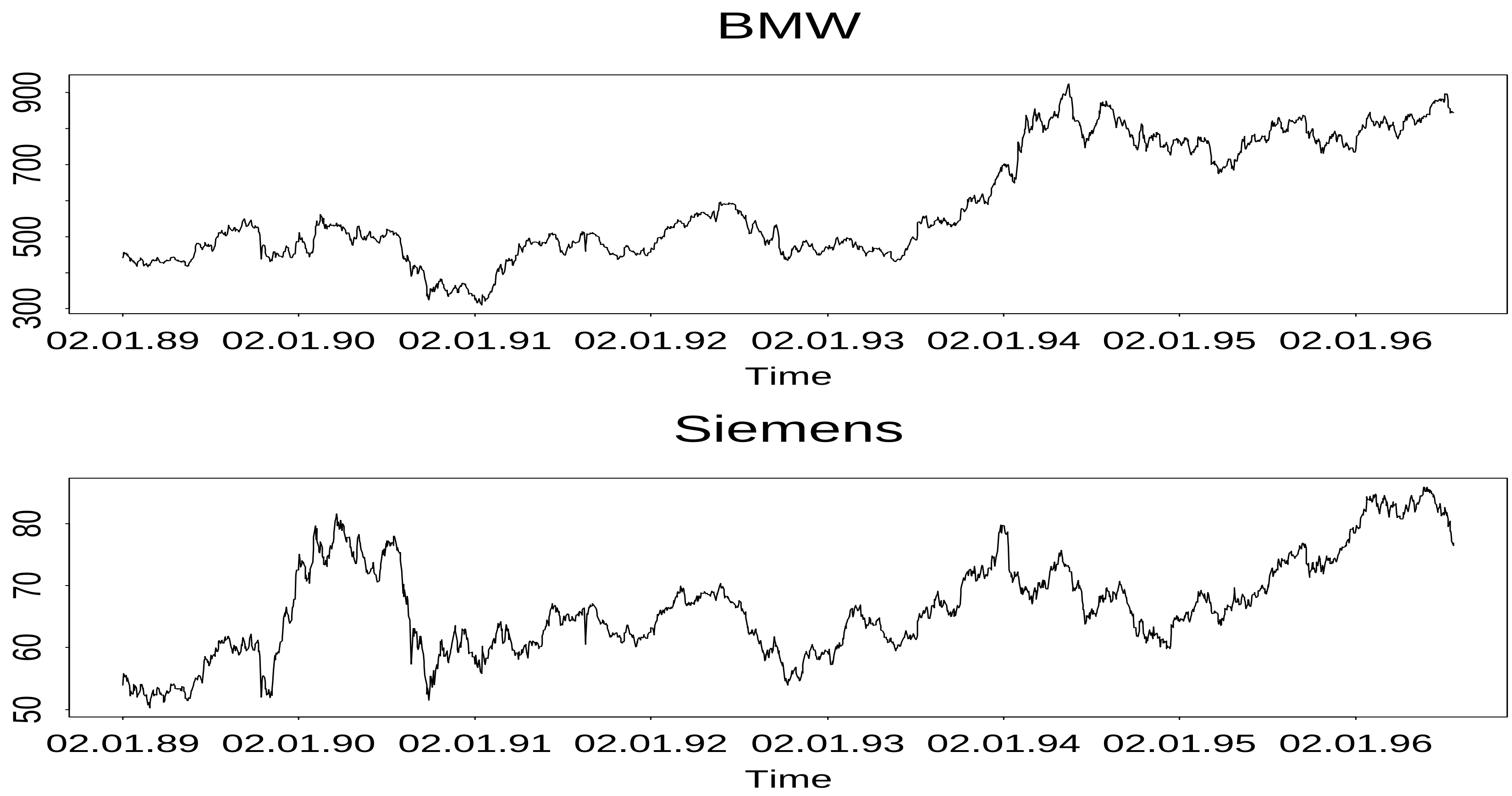
Consider d stocks; let α_i denote number of shares in stock i at time t and let $S_{t,i}$ denote price.

The risk factors: following standard convention we take logarithmic prices as risk factors $Z_{t,i} = \log S_{t,i}, 1 \leq i \leq d$.

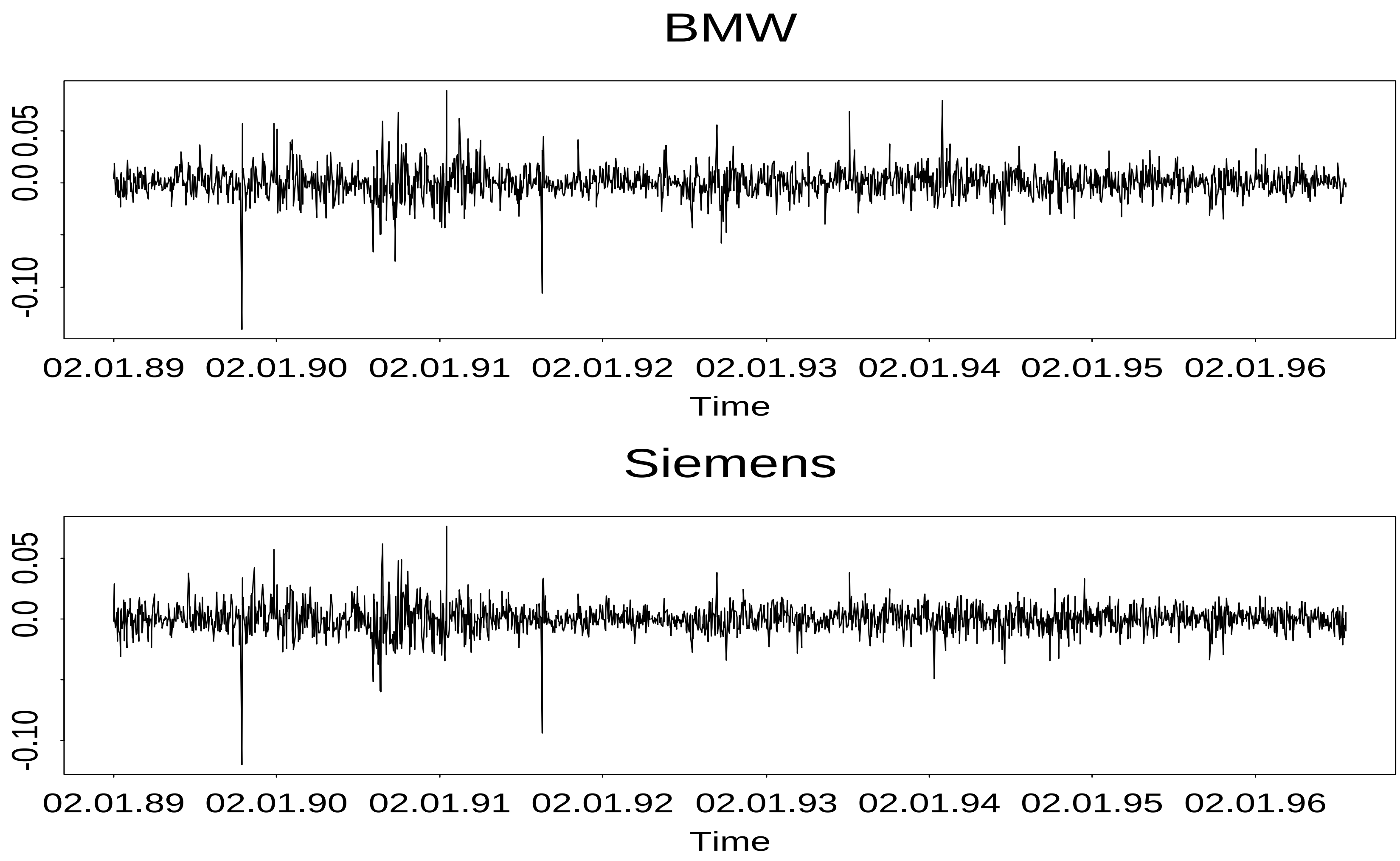
The risk factor changes: in this case these are $X_{t+1,i} = \log S_{t+1,i} - \log S_{t,i}$, which correspond to the so-called log-returns of the stock.

The Mapping (1)

$$V_t = \sum_{i=1}^{d} \alpha_i S_{t,i} = \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}}. \tag{4}$$



BMW and Siemens Data: 1972 days to 23.07.96.
Respective prices on evening 23.07.96: 844.00 and 76.9. Consider portfolio in ratio 1:10 on that evening.



BMW and Siemens Log Return Data: 1972 days to 23.07.96.

Example Continued

The Loss (2)

$$\begin{aligned} L_{t+1} &= -\left(\sum_{i=1}^{d} \alpha_i e^{Z_{t+1,i}} - \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}} \right) \\ &= -V_t \sum_{i=1}^{d} \omega_{t,i} \left(e^{X_{t+1,i}} - 1 \right) \end{aligned} \qquad (5)$$

where $\omega_{t,i} = \alpha_i S_{t,i}/V_t$ is relative weight of stock i at time t.

The loss operator (3)

$$l_{[t]}(\mathbf{x}) = -V_t \sum_{i=1}^{d} \omega_{t,i} \left(e^{x_i} - 1 \right),$$

Numeric Example: $l_{[t]}(\mathbf{x}) = -\left(844(e^{x_1} - 1) + 769(e^{x_2} - 1)\right)$

A3. Conditional or Unconditional Loss Distribution?

This issue is related to the time series properties of $(\mathbf{X}_t)_{t\in\mathbb{N}}$, the series of risk factor changes. If we assume that $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$ are iid random vectors, the issue does not arise. But, if we assume that they form a strictly stationary multivariate time series then we must differentiate between conditional and unconditional.

Many standard accounts of risk management fail to make the distinction between the two.

If we cannot assume that risk factor changes form a stationary time series for at least some window of time extending from the present back into intermediate past, then any statistical analysis of loss distribution is difficult.

The Conditional Problem

Let $\mathcal{F}_t$ represent the history of the risk factors up to the present.

More formally $\mathcal{F}_t$ is sigma algebra generated by past and present risk factor changes $(\mathbf{X}_s)_{s \leq t}$.

In the conditional problem we are interested in the distribution of $L_{t+1} = l_{[t]}(\mathbf{X}_{t+1})$ given $\mathcal{F}_t$, i.e. the conditional (or predictive) loss distribution for the next time interval given the history of risk factor developments up to present.

This problem forces us to model the dynamics of the risk factor time series and to be concerned in particular with predicting volatility. This seems the most suitable approach to market risk.

The Unconditional Problem

In the unconditional problem we are interested in the distribution of $L_{t+1} = l_{[t]}(\mathbf{X})$ when $\mathbf{X}$ is a generic vector of risk factor changes with the same distribution $F_{\mathbf{X}}$ as $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$.

When we neglect the modelling of dynamics we inevitably take this view. Particularly when the time interval is large, it may make sense to do this. Unconditional approach also typical in credit risk.

More Formally

Conditional loss distribution: distribution of $l_{[t]}(\cdot)$ under $F_{[\mathbf{X}_{t+1}|\mathcal{F}_t]}$.

Unconditional loss distribution: distribution of $l_{[t]}(\cdot)$ under $F_{\mathbf{X}}$.

A4. Risk Measures Based on Loss Distributions

Risk measures attempt to quantify the riskiness of a portfolio. The most popular risk measures like VaR describe the right tail of the loss distribution of L_{t+1} (or the left tail of the P&L).

To address this question we put aside the question of whether to look at conditional or unconditional loss distribution and assume that this has been decided.

Denote the distribution function of the loss $L := L_{t+1}$ by F_L so that $P(L \leq x) = F_L(x)$.

VaR and Expected Shortfall

- Primary risk measure: Value at Risk defined as

$$\mathrm{VaR}_\alpha = q_\alpha(F_L) = F_L^{\leftarrow}(\alpha)\,, \tag{6}$$

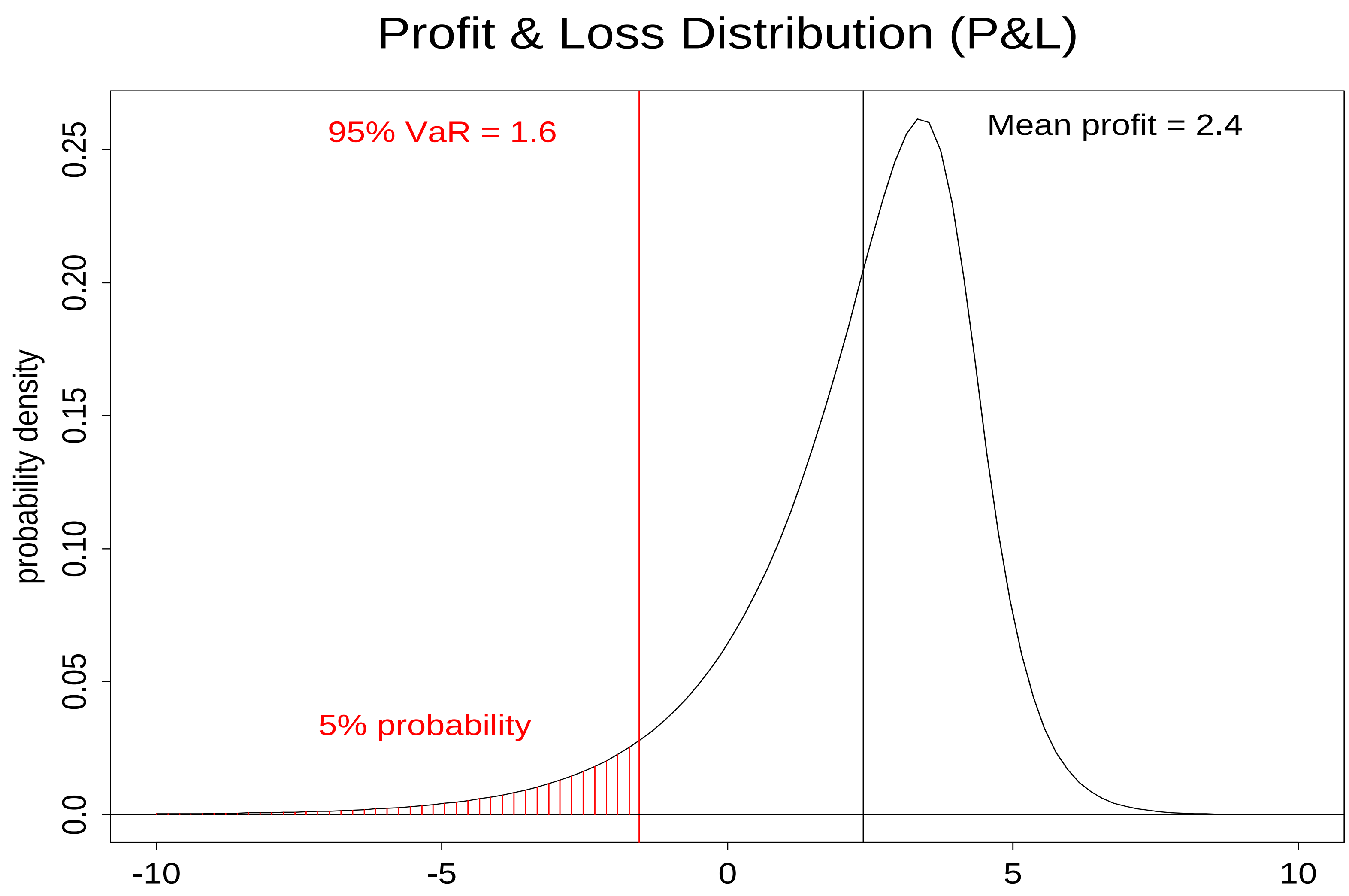
  i.e. the α-quantile of F_L.

- Alternative risk measure: Expected shortfall defined as

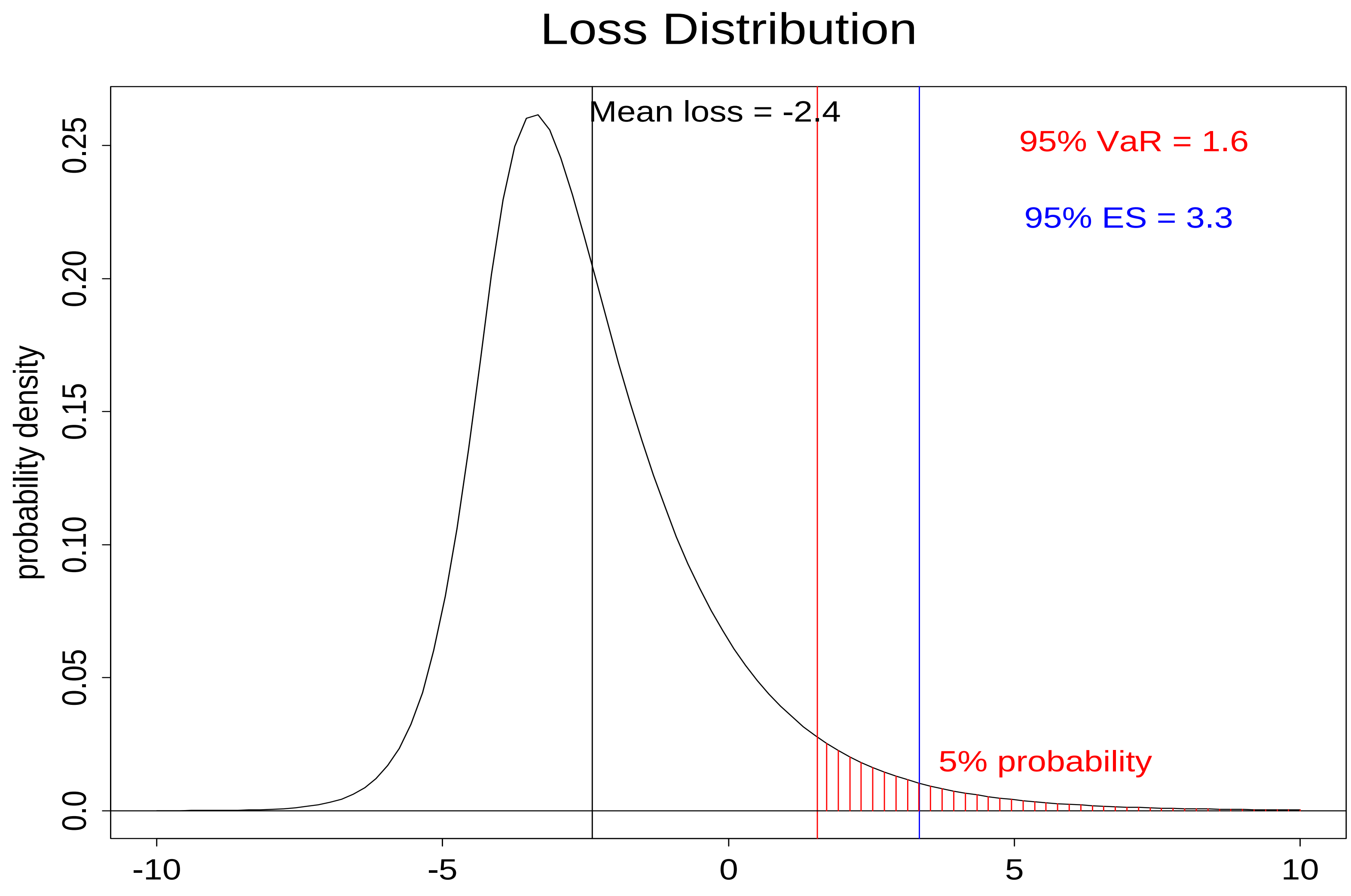
$$\mathsf{ES}_\alpha = E\left(L \mid L > \mathrm{VaR}_\alpha\right)\,, \tag{7}$$

  i.e. the average loss when VaR is exceeded. ES_α gives information about frequency and size of large losses.

VaR in Visual Terms



Losses and Profits



VaR - badly defined!

The VaR bible is the book by Philippe Jorion.[Jorion, 2001].

The following "definition" is very common:

"VaR is the $maximum$ expected loss of a portfolio over a given time horizon with a certain confidence level."

It is however mathematically meaningless and potentially misleading.
In no sense is VaR a maximum loss!
We can lose more, sometimes much more, depending on the heaviness of the tail of the loss distribution.

A5. Linearisation of Loss

Recall the general formula (2) for the loss L_{t+1} in time period $[t, t+1]$. If the mapping f is differentiable we may use the following first order approximation for the loss

$$L^{\Delta}_{t+1} = -\left(f_t(t, \mathbf{Z}_t) + \sum_{i=1}^{d} f_{z_i}(t, \mathbf{Z}_t) X_{t+1,i} \right), \qquad (8)$$

- ⋆ f_{z_i} is partial derivative of mapping with respect to risk factor i
  ⋆ f_t is partial derivative of mapping with respect to time
- The term $f_t(t, \mathbf{Z}_t)$ only appears when mapping explicitly features time (derivative portfolios) and is sometimes neglected.

Linearised Loss Operator

Recall the loss operator (3) which applies at time t. We can obviously also define a linearised loss operator

$$l_{[t]}^{\Delta}(\mathbf{x}) = -\left(f_t(t, \mathbf{z}_t) + \sum_{i=1}^{d} f_{z_i}(t, \mathbf{z}_t) x_i \right), \tag{9}$$

where notation is as in previous slide and $\mathbf{z}_t$ is realisation of $\mathbf{Z}_t$.

Linearisation is convenient because linear functions of the risk factor changes may be easier to handle analytically. It is crucial to the variance-covariance method. The quality of approximation is best if we are measuring risk over a short time horizon and if portfolio value is almost linear in risk factor changes.

Stock Portfolio Example

Here there is no explicit time dependence in the mapping (4). The partial derivatives with respect to risk factors are

$$f_{z_i}(t, \mathbf{z}_t) = \alpha_i e^{z_{t,i}}, \ 1 \leq i \leq d,$$

and hence the linearised loss (8) is

$$L^{\Delta}_{t+1} = -\sum_{i=1}^{d} \alpha_i e^{\mathbf{Z}_{t,i}} X_{t+1,i} = -V_t \sum_{i=1}^{d} \omega_{t,i} X_{t+1,i},$$

where $\omega_{t,i} = \alpha_i S_{t,i}/V_t$ is relative weight of stock i at time t. This formula may be compared with (5).

Numeric Example: $l^{\Delta}_{[t]}(\mathbf{x}) = -\left(844x_1 + 769x_2\right)$

A6. Example: European Call Option

Consider portfolio consisting of one standard European call on a non-dividend paying stock S with maturity T and exercise price K.

The Black-Scholes value of this asset at time t is $C^{BS}(t, S_t, r, \sigma)$ where

$$C^{BS}(t, S; r, \sigma) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

Φ is standard normal df, r represents risk-free interest rate, σ the volatility of underlying stock, and where

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \text{ and } d_2 = d_1 - \sigma\sqrt{T-t}.$$

While in BS model, it is assumed that interest rates and volatilities are constant, in reality they tend to fluctuate over time; they should be added to our set of risk factors.

The Issue of Time Scale

Rather than measuring time in units of the time horizon (as we have implicitly done in most of this chapter) it is more common when derivatives are involved to measure time in years (as in the Black Scholes formula).

If Δ is the length of the time horizon measured in years (i.e. $\Delta = 1/260$ if time horizon is one day) then we have

$$V_t = f(t, \mathbf{Z}_t) = C^{BS}(t\Delta, S_t; r_t, \sigma_t).$$

When linearising we have to recall that

$$f_t(t, \mathbf{Z}_t) = C_t^{BS}(t\Delta, S_t; r_t, \sigma_t)\Delta.$$

Example Summarised

The risk factors: $\mathbf{Z}_t = (\log S_t, r_t, \sigma_t)'$.

The risk factor changes:
$\mathbf{X}_t = (\log(S_t/S_{t-1}), r_t - r_{t-1}, \sigma_t - \sigma_{t-1})'$.

The mapping (1)

$$V_t = f(t, \mathbf{Z}_t) = C^{BS}(t\Delta, S_t; r_t, \sigma_t),$$

The loss/loss operator could be calculated from (2). For derivative positions it is quite common to calculate linearised loss.

The linearised loss (8)

$$L_{t+1}^{\Delta} = -\left(f_t(t, \mathbf{Z}_t) + \sum_{i=1}^{3} f_{z_i}(t, \mathbf{Z}_t) X_{t+1,i} \right).$$

The Greeks

It is more common to write the linearised loss as

$$L_{t+1}^{\Delta} = -\left(C_t^{BS}\Delta + C_S^{BS}S_tX_{t+1,1} + C_r^{BS}X_{t+1,2} + C_\sigma^{BS}X_{t+1,3}\right),$$

in terms of the derivatives of the BS formula.

- C_S^{BS} is known as the delta of the option.
- C_σ^{BS} is the vega.
- C_r^{BS} is the rho.
- C_t^{BS} is the theta.

References

On risk management:

- [McNeil et al., 2004] (methods for QRM)
- [Crouhy et al., 2001] (on risk management)
- [Jorion, 2001] (on VaR)
- [Artzner et al., 1999] (coherent risk measures)

B. Standard Statistical Methods for Market Risk

1. Variance-Covariance Method

2. Historical Simulation Method

3. Monte Carlo Simulation Method

4. An Example

5. Improving the Statistical Toolkit

B1. Variance-Covariance Method

Further Assumptions

- We assume $\mathbf{X}_{t+1}$ has a multivariate normal distribution (either unconditionally or conditionally).

- We assume that the linearized loss in terms of risk factors is a sufficiently accurate approximation of the loss. We consider the problem of estimating the distribution of

$$L^{\Delta} = l^{\Delta}_{[t]}(\mathbf{X_{t+1}}),$$

Theory Behind Method

Assume $\mathbf{X}_{t+1} \sim N_d(\boldsymbol{\mu}, \Sigma)$.

Assume the linearized loss operator (9) has been determined and write this for convenience as

$$l^{\Delta}_{[t]}(\mathbf{x}) = -\left(c + \sum_{i=1}^{d} w_i x_i\right) = -(c + \mathbf{w}'\mathbf{x}).$$

The loss distribution is approximated by the distribution of $L^{\Delta} = l^{\Delta}_{[t]}(\mathbf{X_{t+1}})$.

Now since $\mathbf{X}_{t+1} \sim N_d(\boldsymbol{\mu}, \Sigma) \Rightarrow \mathbf{w}'\mathbf{X}_{t+1} \sim N(\mathbf{w}'\boldsymbol{\mu}, \mathbf{w}'\Sigma\mathbf{w})$, we have

$$L^{\Delta} \sim N(-c - \mathbf{w}'\boldsymbol{\mu}, \mathbf{w}'\Sigma\mathbf{w}).$$

Implementing the Method

1. The constant terms in c and $\mathbf{w}$ are calculated

2. The mean vector $\boldsymbol{\mu}$ and covariance matrix Σ are estimated from data $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$ to give estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\Sigma}$.

3. Inference about the loss distribution is made using distribution $N(-c - \mathbf{w}'\widehat{\boldsymbol{\mu}}, \mathbf{w}'\widehat{\Sigma}\mathbf{w})$

4. Estimates of the risk measures VaR_α and ES_α are calculated from the estimayed distribution of L^Δ.

Estimating Risk Measures

- Value-at-Risk. VaR_α is estimated by

$$\widehat{\mathrm{VaR}}_\alpha = -c - \mathbf{w}'\widehat{\boldsymbol{\mu}} + \sqrt{\mathbf{w}'\widehat{\Sigma}\mathbf{w}} \cdot \Phi^{-1}(\alpha).$$

- Expected Shortfall. ES_α is estimated by

$$\widehat{ES}_\alpha = -c - \mathbf{w}'\widehat{\boldsymbol{\mu}} + \sqrt{\mathbf{w}'\widehat{\Sigma}\mathbf{w}} \cdot \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$$

Remark. For a rv $Y \sim N(0,1)$ it can be shown that
$E(Y \mid Y > \Phi^{-1}(\alpha)) = \phi(\Phi^{-1}(\alpha))/(1-\alpha)$
where ϕ is standard normal density and Φ the df.

Pros and Cons, Extensions

- Pros. In contrast to the methods that follow, variance-covariance offers analytical solution with no simulation.

- Cons. Linearization may be crude approximation. Assumption of normality may seriously underestimate tail of loss distribution.

- Extensions. Instead of assuming normal risk factors, the method could be easily adapted to use multivariate Student t risk factors or multivariate hyperbolic risk factors, without sacrificing tractibility. (Method works for all elliptical distributions.)

B2. Historical Simulation Method

The Idea

Instead of estimating the distribution of $L = l_{[t]}(\mathbf{X}_{t+1})$ under some explicit parametric model for $\mathbf{X}_{t+1}$, estimate distribution of the loss operator under empirical distribution of data $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$.

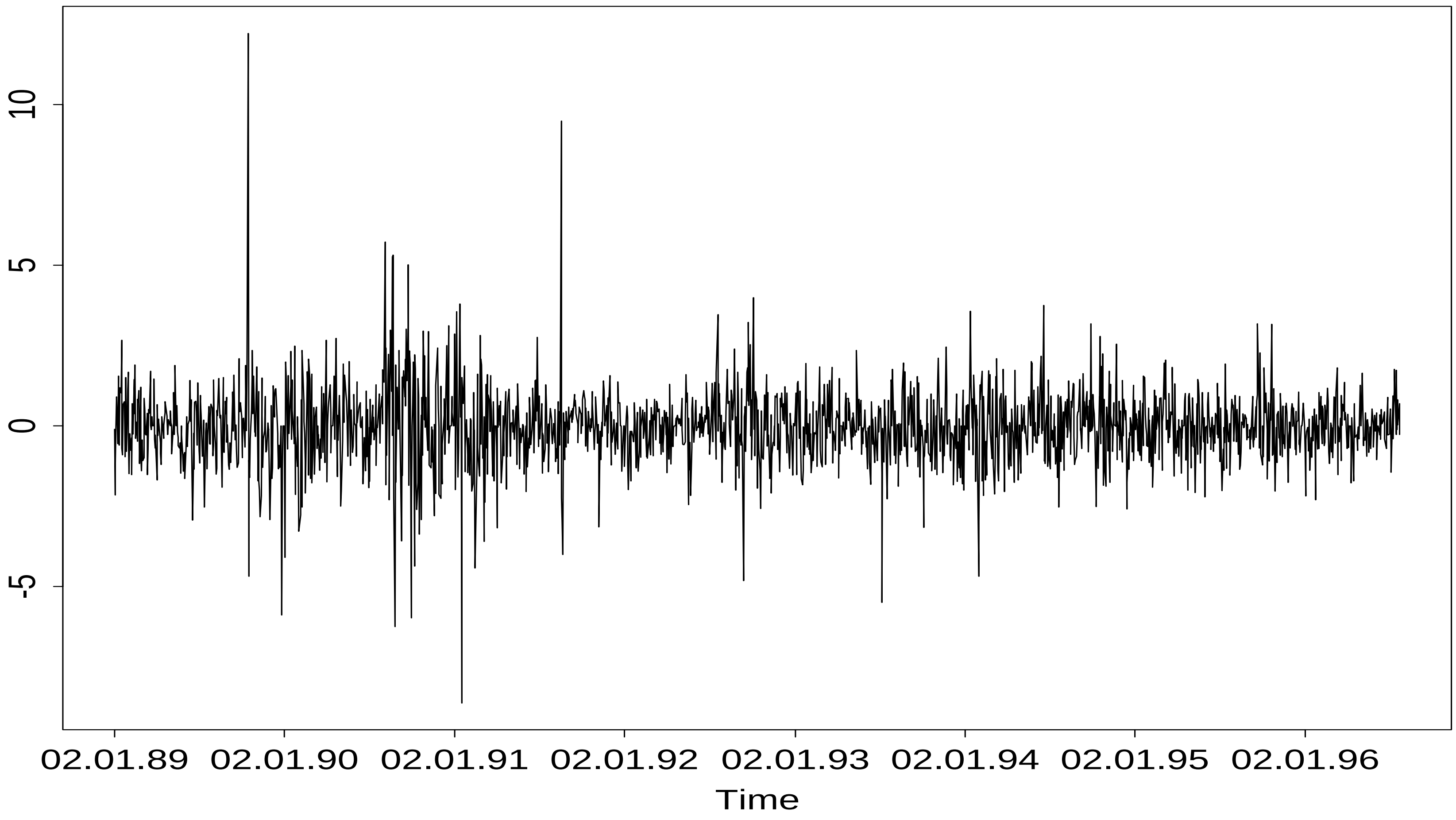
The Method

1. Construct the historical simulation data

$$\{\widetilde{L}_s = l_{[t]}(\mathbf{X}_s) \, : \, s = t - n + 1, \ldots, t\} \tag{10}$$

2. Make inference about loss distribution and risk measures using these historically simulated data: $\widetilde{L}_{t-n+1}, \ldots, \widetilde{L}_t$.

Historical Simulation Data: Percentage Losses



Inference about loss distribution

There are various possibilities in a simulation approach:

- Use empirical quantile estimation to estimate the VaR directly from the simulated data. But what about precision?
- Fit a parametric univariate distribution to $\widetilde{L}_{t-n+1}, \ldots, \widetilde{L}_t$ and calculate risk measures from this distribution.
  But which distribution, and will it model the tail?
- Use the techniques of extreme value theory to estimate the tail of the loss distribution and related risk measures.

Theoretical Justification

If $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$ are iid or more generally stationary, convergence of empirical distribution to true distribution is ensured by suitable version of law of large numbers.

Pros and Cons

- Pros. Easy to implement. No statistical estimation of the distribution of $\mathbf{X}$ necessary.
- Cons. It may be difficult to collect sufficient quantities of relevant, synchronized data for all risk factors. Historical data may not contain examples of extreme scenarios.

B3. The Monte Carlo Method

Idea

We estimate the distribution of $L = l_{[t]}(\mathbf{X}_{t+1})$ under some explicit parametric model for $\mathbf{X}_{t+1}$.

In contrast to the variance-covariance approach we do not necessarily make the problem analytically tractible by linearizing the loss and making an assumption of normality for the risk factors.

Instead we make inference about L using Monte Carlo methods, which involves simulation of new risk factor data.

The Method

1. With the help of the historical risk factor data $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$ calibrate a suitable statistical model for risk factor changes and simulate m new data $\widetilde{\mathbf{X}}_{t+1}^{(1)}, \ldots, \widetilde{\mathbf{X}}_{t+1}^{(m)}$ from this model.

2. Construct the Monte Carlo data $\{\widetilde{L}_i = l_{[t]}(\widetilde{\mathbf{X}}_{t+i}^{(i)}),\ i = 1, \ldots, m\}$.

3. Make inference anout loss distribution and risk measures using the simulated data $\widetilde{L}_1, \ldots, \widetilde{L}_m$. We have similar possibilities as for historical simulation.

Pros and Cons

- Pros. Very general. No restriction in our choice of distribution for $\mathbf{X}_{t+1}$.

- Cons. Can be very time consuming if loss operator is difficult to evaluate, which depends on size and complexity of portfolio.

Note that MC approach does not address the problem of determining the distribution of $\mathbf{X}_{t+1}$.

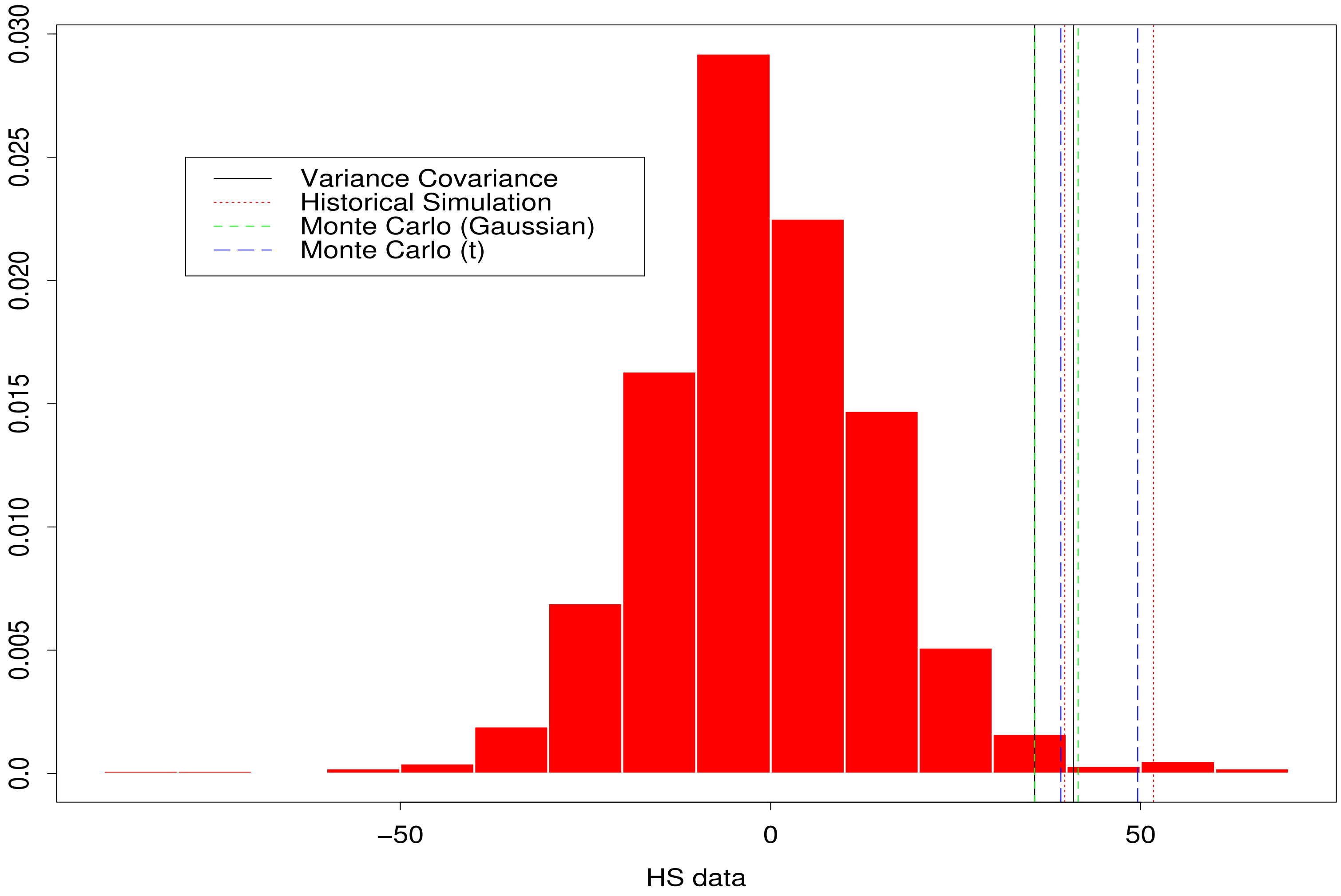
B4. An Example With BMW-SIEMENS Data

```
> Xdata <- DAX[(5147:6146),c("BMW","SIEMENS")]
> X <- seriesData(Xdata)

# Set stock prices and number of units
> alpha <- cbind(1,10)
> Sprice <- cbind(844,76.9)

#1. Implement variance-covariance analysis

> weights <- alpha*Sprice
> muhat <- apply(X,2,mean)
> Sigmahat  <- var(X)
> meanloss <- -sum(weights*muhat)
> varloss <- weights %*% Sigmahat %*% t(weights)
> VaR99 <- meanloss + sqrt(varloss)*qnorm(0.99)
> ES99 <- meanloss +sqrt(varloss)*dnorm(qnorm(0.99))/0.01

#2. Implement a historical simulation analysis

> loss.operator <- function(x,weights){
-apply((exp(x)-1)*matrix(weights,nrow=dim(x)[1],ncol=length(weights),byrow=T),1,sum)}
> hsdata <- loss.operator(X,weights)
> VaR99.hs <- quantile(hsdata,0.99)
> ES99.hs <- mean(hsdata[hsdata > VaR99.hs])
```

Example Continued

```
#3a. Implement a Monte Carlo simulation analysis with Gaussian risk factors

> X.new <- rmnorm(10000,Sigma=Sigmahat,mu=muhat)
> mcdata <- loss.operator(X.new,weights)
> VaR99.mc <- quantile(mcdata,0.99)
> ES99.mc <- mean(mcdata[mcdata > VaR99.mc])

#3b. Implement alternative Monte Carlo simulation analysis with t risk factors

> model <- fit.t(X, nu=NA)
> X.new <- rmt(10000, df=model$nu, Sigma=model$Sigma, mu=model$mu)
> mcdatat <- loss.operator(X.new,weights)
> VaR99.mct <- quantile(mcdatat,0.99)
> ES99.mct <- mean(mcdatat[mcdatat > VaR99.mct])

#Draw pictures

> hist(hsdata,nclass=20,prob=T)
> abline(v=c(VaR99,ES99))
> abline(v=c(VaR99.hs,ES99.hs),col=2)
> abline(v=c(VaR99.mc,ES99.mc),col=3)
> abline(v=c(VaR99.mct,ES99.mct),col=4)
```

Comparison of Risk Measure Estimates



B5. Improving the Statistical Toolkit

Questions we will examine in the remainder of this workshop include the following.

Multivariate Models

Are there alternatives to the multivariate normal distribution for modelling changes in several risk factors?

We will expand our stock of multivariate models to include multivariate normal mixture models and copula models. These will allow a more realistic description of joint extreme risk factor changes.

Improving the Statistical Toolkit II

Monte Carlo Techniques

How can we simulate dependent risk factor changes?

We will look in particular at ways of simulating multivariate risk factors in non-Gaussian models.

Conditional Risk Measurement

How can we implement a genuinely conditional calculation of risk measures that takes the dynamics of risk factors into consideration?

We will consider methodology for modelling financial time series and predicting volatility, particularly using GARCH models.

References

On risk management:

- [Crouhy et al., 2001]
- [Jorion, 2001]

C. Fundamentals of Modelling Dependent Risks

1. Motivation: Multivariate Risk Factor Data

2. Basics of Multivariate Statistics

3. The Multivariate Normal Distribution

4. Standard Estimators of Location and Dispersion

5. Tests of Multivariate Normality

6. Dimension Reduction and Factor Models

C1. Motivation: Multivariate Risk Factor Data

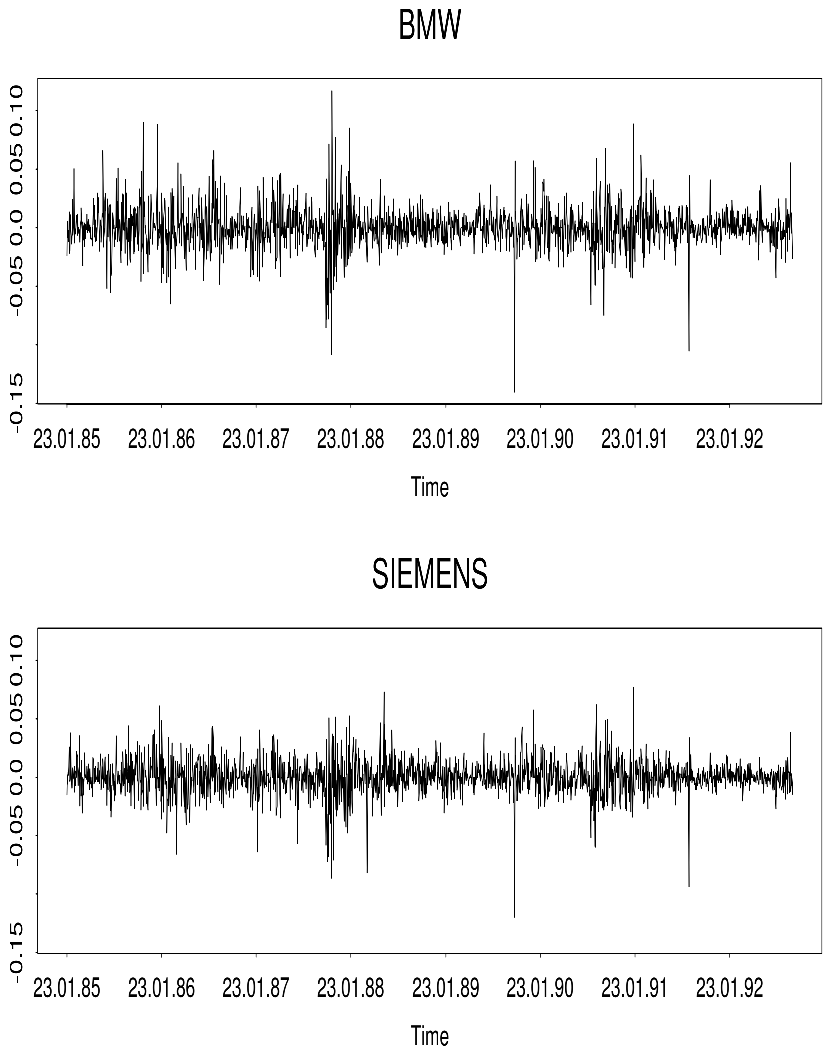
Assume we have data on risk factor changes $\mathbf{X}_1, \ldots, \mathbf{X}_n$. These might be daily (log) returns in context of market risk or longer interval returns in credit risk (e.g. monthly/yearly asset value returns). What are appropriate multivariate models?

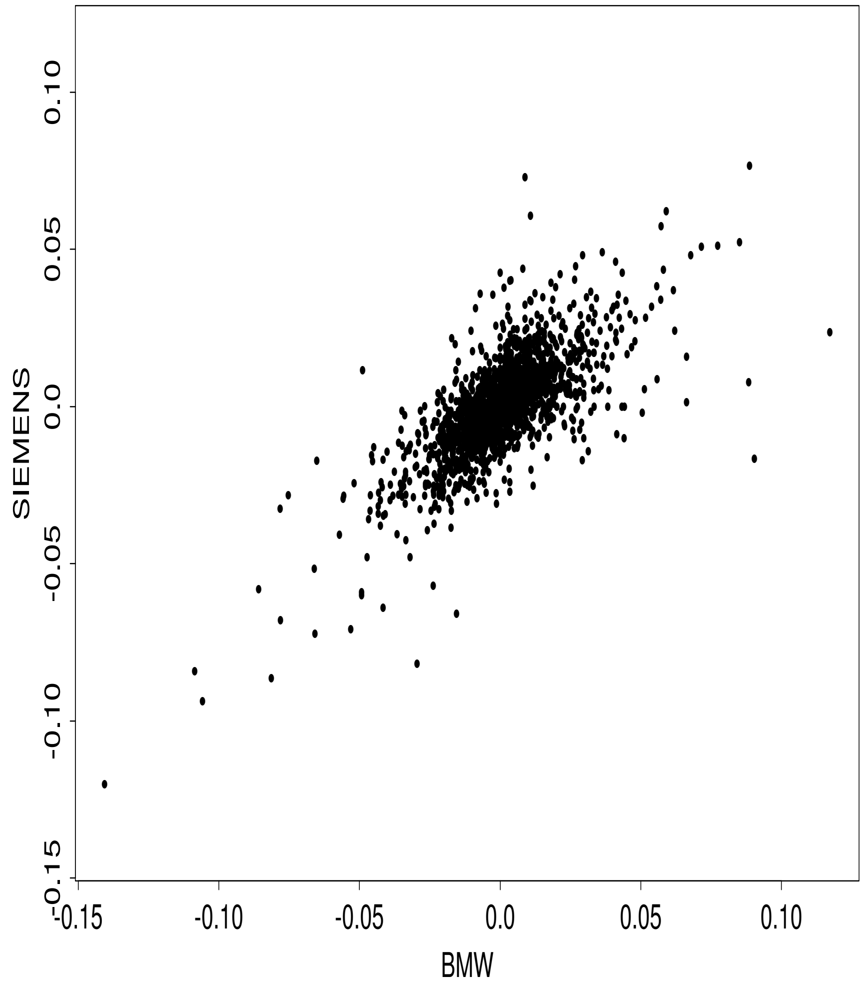
- Distributional Models. In unconditional approach to risk modelling we require appropriate multivariate distributions, which are calibrated under assumption data come from stationary time series.

- Dynamic Models. In conditional approach we use multivariate time series models that allow us to make risk forecasts.

This module concerns the first issue. A motivating example shows the kind of data features that particularly interest us.

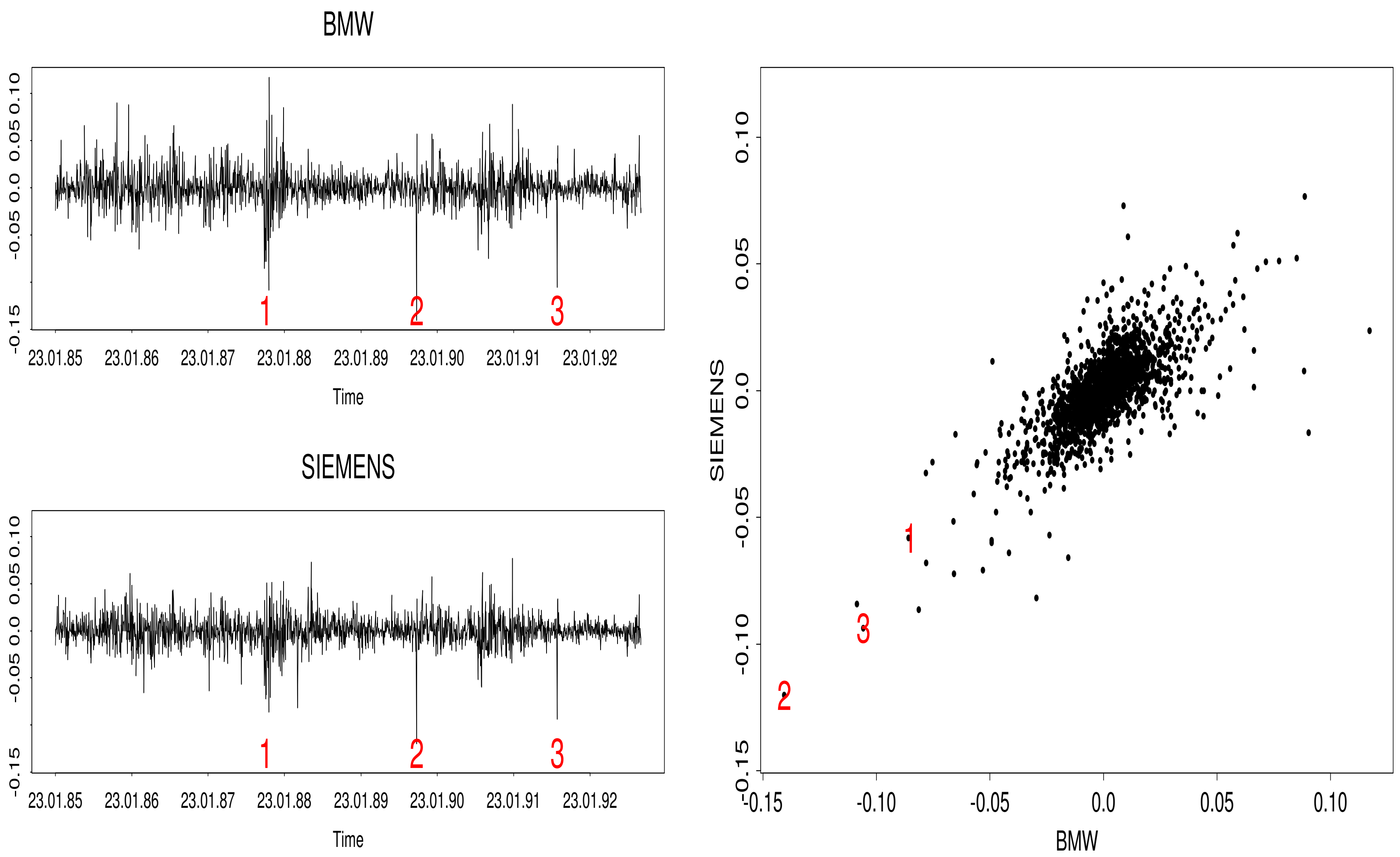
Bivariate Daily Return Data





BMW and Siemens: 2000 daily (log) returns 1985-1993.

Three Extreme Days



Those extreme days: 19.10.1987, 16.10.1989, 19.08.1991

History


The Philadelphia Inquirer

Dow Dives 508.32 Point
In Panic on Wall Stree

Navy blasts
Iran oil site
in retaliation

Billio
lost in
tradin

Archive Photos/Ed Carlin

New York, 19th October 1987

Berlin Wall 

16thOctober 1989

The Kremlin, 19th August 1991 

C2. Multivariate Statistics: Basics

Let $\mathbf{X} = (X_1, \ldots, X_d)'$ be a d-dimensional random vector representing risks of various kinds. Possible interpretations:

- returns on d financial instruments (market risk)
- asset value returns for d companies (credit risk)
- results for d lines of business (risk integration)

An individual risk X_i has marginal df $F_i(x) = P(X_i \leq x)$. A random vector of risks has joint df

$$F(\mathbf{x}) = F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d)$$

or joint survivor function

$$\overline{F}(\mathbf{x}) = \overline{F}(x_1, \ldots, x_d) = P(X_1 > x_1, \ldots, X_d > x_d).$$

Multivariate Models

If we fix F (or $\overline{F}$) we specify a multivariate model and implicitly describe marginal behaviour and dependence structure of the risks.

Calculating Marginal Distributions

$$F_i(x_i) = P\left(X_i \le x_i\right) = F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty),$$

i.e. limit as arguments tend to infinity.

In a similar way higher dimensional marginal distributions can be calculated for other subsets of $\{X_1, \ldots, X_d\}$.

Independence

$X_1, \ldots, X_d$ are said to be mutually independent if

$$F(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i), \quad \forall \mathbf{x} \in \mathbb{R}^d.$$

Densities of Multivariate Distributions

Most, but not all, of the models we consider can also be described by joint densities $f(\mathbf{x}) = f(x_1, \ldots, x_d)$, which are related to the joint df by

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

Existence of a joint density implies existence of marginal densities $f_1, \ldots, f_d$ (but not vice versa).

Equivalent Condition for Independence

$$f(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i), \quad \forall \mathbf{x} \in \mathbb{R}^d$$

C3. Multivariate Normal (Gaussian) Distribution

This distribution can be defined by its density

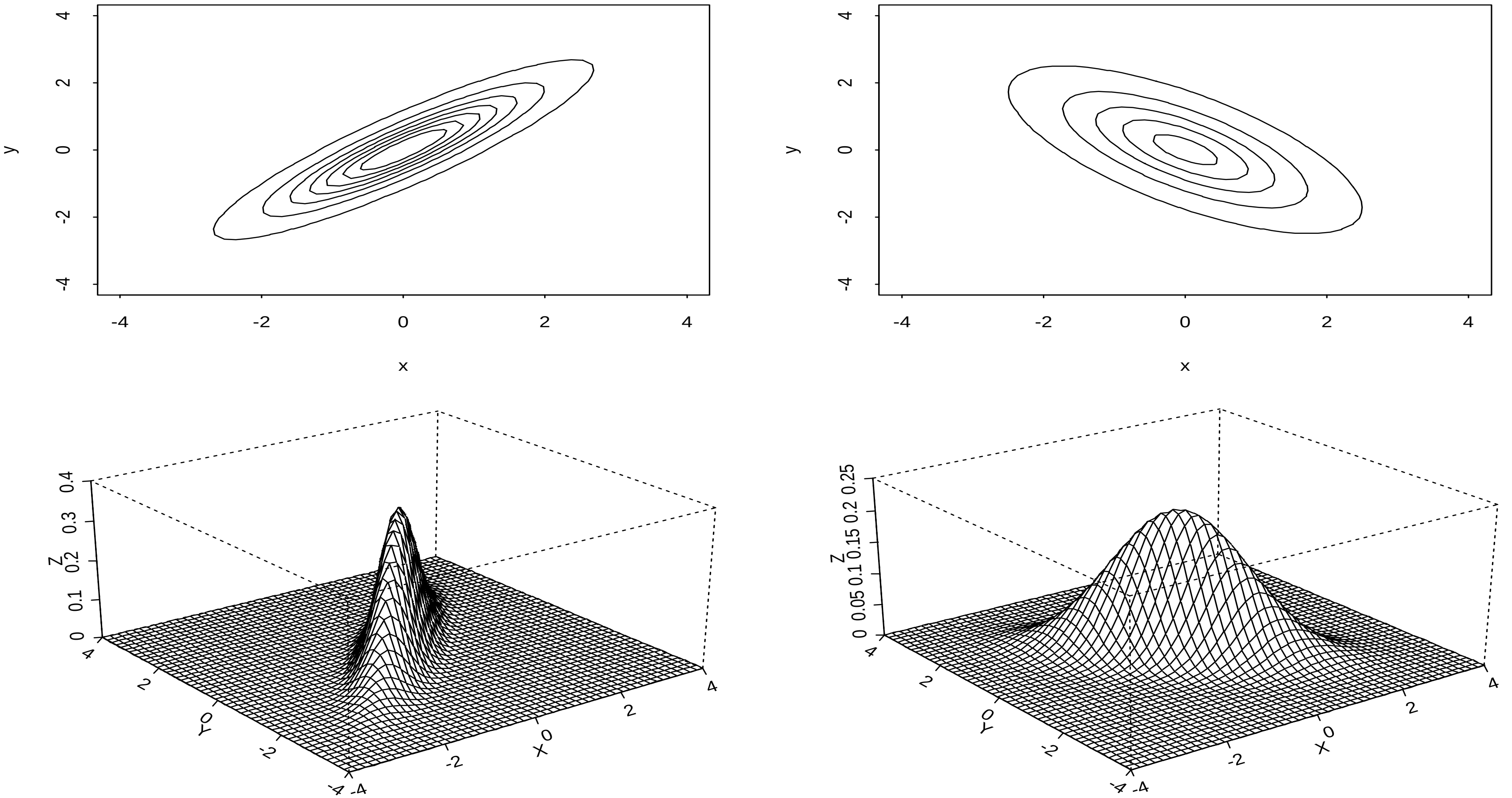
$$f(\mathbf{x}) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right\},$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix.

- If $\mathbf{X}$ has density f then $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\operatorname{cov}(\mathbf{X}) = \Sigma$, so that $\boldsymbol{\mu}$ and Σ are the mean vector and covariance matrix respectively. A standard notation is $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$.

- Clearly, the components of $\mathbf{X}$ are mutually independent if and only if Σ is diagonal. For example, $\mathbf{X} \sim N_d(\mathbf{0}, I)$ if and only if $X_1, \ldots, X_d$ are iid $N(0,1)$.

Bivariate Standard Normals



In left plots $\rho = 0.9$; in right plots$\rho = -0.7$.

Properties of Multivariate Normal Distribution

- The marginal distributions are univariate normal.
- Linear combinations $\mathbf{a}'\mathbf{X} = a_1 X_1 + \cdots a_d X_d$ are univariate normal with distribution $\mathbf{a}'\mathbf{X} \sim N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a})$.
- Conditional distributions are multivariate normal.
- The sum of squares $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2_d$ (chi-squared).

Simulation.

1. Perform a Cholesky decomposition $\Sigma = AA'$
2. Simulate iid standard normal variates $\mathbf{Z} = (Z_1, \ldots, Z_d)'$
3. Set $\mathbf{X} = \boldsymbol{\mu} + A\mathbf{Z}$.

C4. Estimators of Location and Dispersion

Assumptions. We have data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ which are either iid or at least serially uncorrelated from a distribution with mean vector $\boldsymbol{\mu}$, finite covariance matrix Σ and correlation matrix P.

Standard method-of-moments estimators of $\boldsymbol{\mu}$ and Σ are the sample mean vector $\overline{\mathbf{X}}$ and the sample covariance matrix S defined by

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i, \quad S = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'.$$

These are unbiased estimators.

The sample correlation matrix has (i,j)th element given by $R_{ij} = S_{ij}/\sqrt{S_{ii}S_{jj}}$. Defining D to be a d-dimensional diagonal matrix with ith diagonal element $S_{ii}^{1/2}$ we may write $R = D^{-1}SD^{-1}$.

Properties of the Estimators?

Further properties of the estimators $\overline{\mathbf{X}}$, S and R depend on the true multivariate distribution of observations. They are not necessarily the best estimators of $\boldsymbol{\mu}$, Σ and P in all situations, a point that is often forgotten in financial risk management where they are routinely used.

If our data are iid multivariate normal $N_d(\boldsymbol{\mu}, \Sigma)$ then $\overline{\mathbf{X}}$ and $(n-1)S/n$ are the maximum likelihood estimators (MLEs) of the mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. Their behaviour as estimators is well understood and statistical inference concerning the model parameters is relatively unproblematic.

However, certainly at short time intervals such as daily data, the multivariate normal is not a good description of financial risk factor returns and other estimators of $\boldsymbol{\mu}$ and Σ may be better.

C5. Testing for Multivariate Normality

If data are to be multivariate normal then margins must be univariate normal. This can be assessed graphically with QQplots or tested formally with tests like Jarque-Bera or Anderson-Darling.

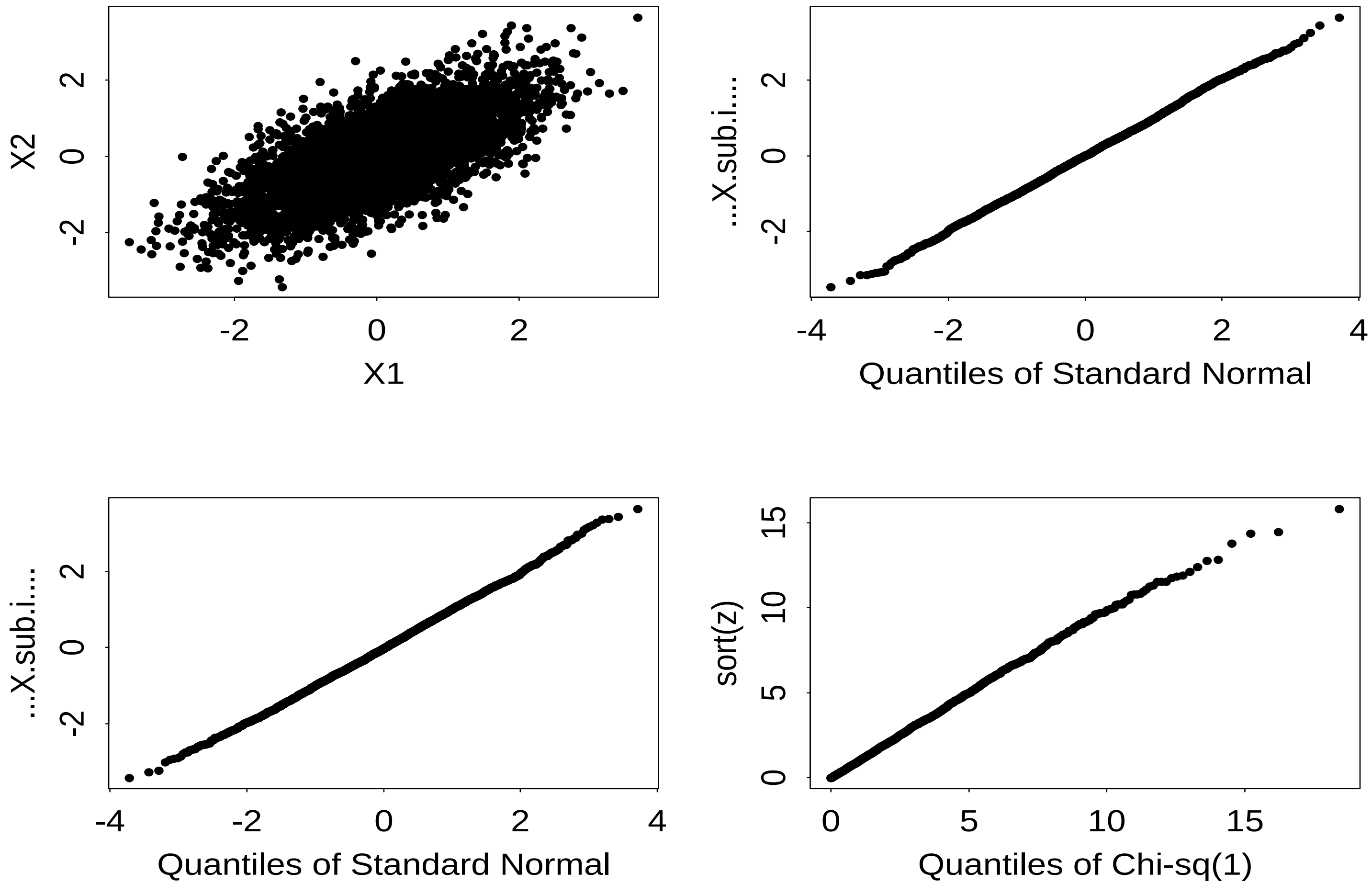
However, normality of the margins is not sufficient – we must test joint normality. To this end we calculate

$$\left\{(\mathbf{X}_i - \widehat{\boldsymbol{\mu}})' \, \widehat{\Sigma}^{-1} \, (\mathbf{X}_i - \widehat{\boldsymbol{\mu}}) \, , \; i = 1, \ldots, n\right\}.$$

These should form (approximately) a sample from a χ^2_d–distribution, and this can be assessed with a QQplot or tested numerically with, for example, Kolmogorov-Smirnov.

(QQplots compare empirical quantiles with theoretical quantiles of reference distribution.)

Testing Multivariate Normality: Normal Data



Deficiencies of Multivariate Normal for Risk Factors

- Tails of univariate margins are very thin and generate too few extreme values.

- Simultaneous large values in several margins relatively infrequent. Model cannot capture phenomenon of joint extreme moves in several risk factors.

- Very strong symmetry (known as elliptical symmetry). Reality suggests more skewness present.

C6. Dimension Reduction and Factor Models

Idea: Explain the variability in a d-dimensional vector $\mathbf{X}$ in terms of a smaller set of common factors.

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

$$\mathbf{X} = \mathbf{a} + B\mathbf{F} + \boldsymbol{\varepsilon}, \tag{11}$$

(i) $\mathbf{F} = (F_1, \ldots, F_p)'$ is random vector of factors with $p < d$,
(ii) $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)'$ is random vector of idiosyncratic error terms, which are uncorrelated and mean zero,
(iii) $B \in \mathbb{R}^{d \times p}$ is a matrix of constant factor loadings and $\mathbf{a} \in \mathbb{R}^d$ a vector of constants,
(iv) $\mathrm{cov}(\mathbf{F}, \boldsymbol{\varepsilon}) = E((\mathbf{F} - E(\mathbf{F}))\boldsymbol{\varepsilon}') = 0.$

Remarks on Theory of Factor Models

- Factor model (11) implies that covariance matrix $\Sigma = \text{cov}(\mathbf{X})$ satisfies $\Sigma = B\Omega B' + \Psi$, where $\Omega = \text{cov}(\mathbf{F})$ and $\Psi = \text{cov}(\boldsymbol{\varepsilon})$ (diagonal matrix).

- Factors can always be transformed so that they are orthogonal:

$$\Sigma = BB' + \Psi. \tag{12}$$

- Conversely, if (12) holds for covariance matrix Σ of random vector $\mathbf{X}$, then $\mathbf{X}$ follows factor model (11) for some $\mathbf{a}, \mathbf{F}$ and $\boldsymbol{\varepsilon}$.

- If, moreover, $\mathbf{X}$ is Gaussian then $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ may be taken to be independent Gaussian vectors, so that $\boldsymbol{\varepsilon}$ has independent components.

Factor Models in Practice

We have multivariate financial return data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ which are assumed to follow (11). Two situations to be distinguished:

1. Appropriate factor data $\mathbf{F}_1, \ldots, \mathbf{F}_n$ are also observed, for example returns on relevant indices. We have a multivariate regression problem; parameters ($\mathbf{a}$ and B) can be estimated by multivariate least squares.

2. Factor data are not directly observed. We assume data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ identically distributed and calibrate factor model by one of two strategies: statistical factor analysis - we first estimate B and Ψ from (12) and use these to reconstruct $\mathbf{F}_1, \ldots, \mathbf{F}_n$; principal components - we fabricate $\mathbf{F}_1, \ldots, \mathbf{F}_n$ by PCA and estimate B and $\mathbf{a}$ by regression.

References

On general multivariate statistics:

- [Mardia et al., 1979] (general multivariate statistics)
- [Seber, 1984] (multivariate statistics)
- [Kotz et al., 2000] (continuous multivariate distributions)

D. Normal Mixture Models and Elliptical Models

1. Normal Variance Mixtures
2. Normal Mean-Variance Mixtures
3. Generalized Hyperbolic Distributions
4. Elliptical Distributions

D1. Multivariate Normal Mixture Distributions

Multivariate Normal Variance-Mixtures

Let $\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$ and let W be an independent, positive, scalar random variable. Let $\boldsymbol{\mu}$ be any deterministic vector of constants. The vector $\mathbf{X}$ given by

$$\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{Z} \tag{13}$$

is said to have a multivariate normal variance-mixture distribution.

Easy calculations give $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{X}) = E(W)\Sigma$.
Correlation matrices of $\mathbf{X}$ and $\mathbf{Z}$ are identical: $\mathrm{corr}(\mathbf{X}) = \mathrm{corr}(\mathbf{Z})$.

Multivariate normal variance mixtures provide the most useful examples of so-called elliptical distributions.

Examples of Multivariate Normal Variance-Mixtures

2 point mixture

$$W = \begin{cases} k_1 \text{ with probability } p, \\ k_2 \text{ with probability } 1-p \end{cases} \qquad k_1 > 0, k_2 > 0, k_1 \neq k_2.$$

Could be used to model two regimes - ordinary and extreme.

Multivariate t

W has an inverse gamma distribution, $W \sim \text{Ig}(\nu/2, \nu/2)$. This gives multivariate t with ν degrees of freedom. Equivalently $\nu/W \sim \chi^2_\nu$.

Symmetric generalised hyperbolic

W has a GIG (generalised inverse Gaussian) distribution.

The Multivariate t Distribution

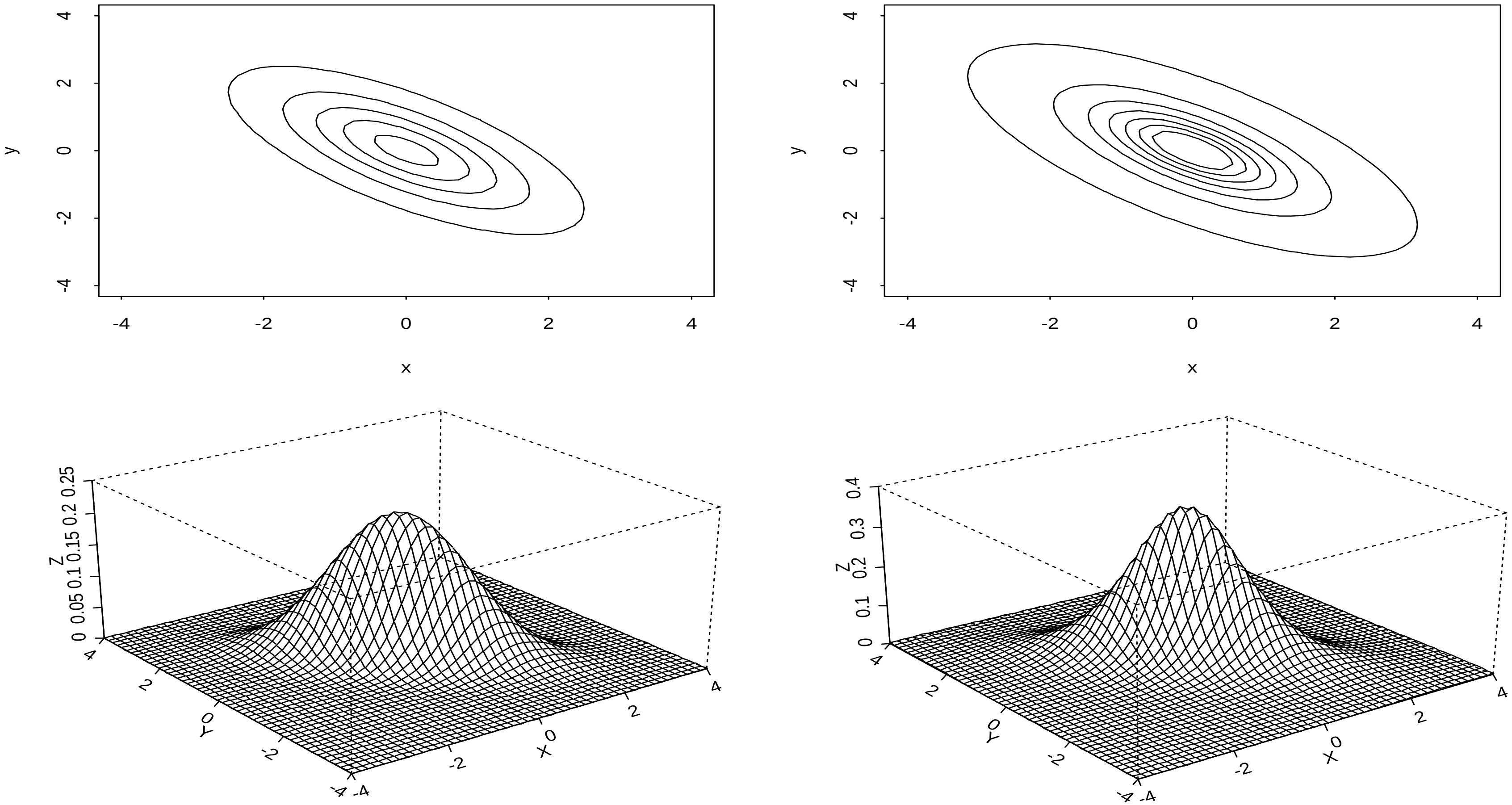
This has density

$$f(\mathbf{x}) = k_{\Sigma,\nu,d}\left(1 + \frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{(\nu+d)}{2}}$$

where $\boldsymbol{\mu} \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times d}$ is a positive definite matrix,ν is the degrees of freedom and $k_{\Sigma,\nu,d}$ is a normalizing constant.

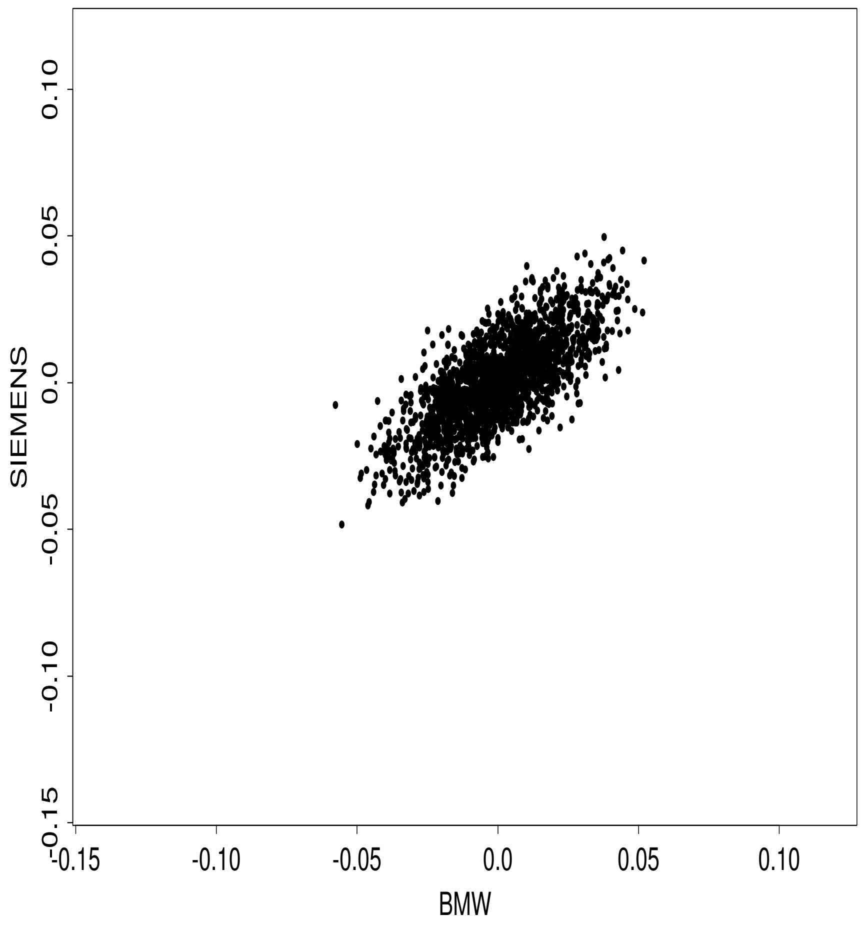
- If $\mathbf{X}$ has density f then $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\operatorname{cov}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$, so that $\boldsymbol{\mu}$ and Σ are the mean vector and dispersion matrix respectively. For finite variances/correlations $\nu > 2$. Notation: $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$.
- If Σ is diagonal the components of $\mathbf{X}$ are uncorrelated. They are not independent.
- The multivariate t distribution has heavy tails.

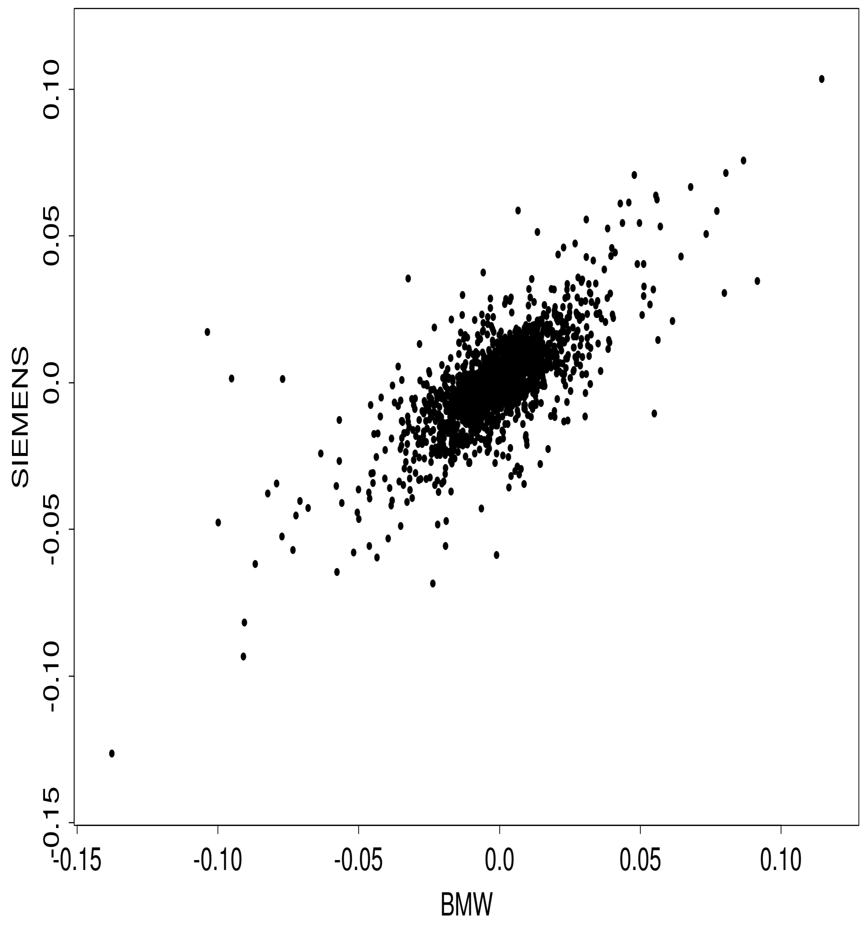
Bivariate Normal and t



$\rho = -0.7$, $\nu = 3$, variances all equal 1.

Fitted Normal and t_3 Distributions





Simulated data (2000) from models fitted by maximum likelihood to BMW-Siemens data.

Simulating Normal–Mixture Distributions

It is straightforward to simulate normal mixture models. We only have to simulate a Gaussian random vector and an independent radial random variable. Simulation of Gaussian vector in all standard texts.

Example: t distribution

To simulate a vector $\mathbf{X}$ with distribution $t_d(\nu, \boldsymbol{\mu}, \Sigma)$ we would simulate $\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$ and $V \sim \chi^2_\nu$;
we would then set $W = \nu/V$ and $\mathbf{X} = \boldsymbol{\mu} + \sqrt{W}\mathbf{Z}$.

To simulate generalized hyperbolic distributions we are required to simulate a radial variate with the GIG distribution. For an algorithm see [Atkinson, 1982]; see also work of [Eberlein et al., 1998].

D2. Multivariate Normal Mean-Variance Mixtures

We can generalise the mixture construction as follows:

$$\mathbf{X} = \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}\mathbf{Z}, \qquad (14)$$

where $\boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^d$ and the positive rv W is again independent of the Gaussian random vector $\mathbf{Z} \sim N_d(\mathbf{0}, \Sigma)$.

This gives us a larger class of distributions, but in general they are no longer elliptical and $\mathrm{corr}(\mathbf{X}) \neq \mathrm{corr}(\mathbf{Z})$. The parameter vector $\boldsymbol{\gamma}$ controls the degree of skewness and $\boldsymbol{\gamma} = \mathbf{0}$ places us back in the (elliptical) variance-mixture family.

Moments of Mean-Variance Mixtures

Since $\mathbf{X} \mid W \sim N_d(\boldsymbol{\mu} + W\boldsymbol{\gamma}, W\Sigma)$ it follows that

$$\begin{aligned} E(\mathbf{X}) &= E\left(E(\mathbf{X} \mid W)\right) = \boldsymbol{\mu} + E(W)\boldsymbol{\gamma}, && (15) \\ \operatorname{cov}(\mathbf{X}) &= E\left(\operatorname{cov}(\mathbf{X} \mid W)\right) + \operatorname{cov}\left(E(\mathbf{X} \mid W)\right) \\ &= E(W)\Sigma + \operatorname{var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}', && (16) \end{aligned}$$

provided W has finite variance. We observe from (15) and (16) that the parameters $\boldsymbol{\mu}$ and Σ are not in general the mean vector and covariance matrix of $\mathbf{X}$.

Note that a finite covariance matrix requires $\operatorname{var}(W) < \infty$ whereas the variance mixtures only require $E(W) < \infty$.

Main example. When W has a GIG distribution we get generalized hyperbolic family.

Generalised Inverse Gaussian (GIG) Distribution

The random variable X has a generalised inverse Gaussian (GIG), written $X \sim N^{-}(\lambda, \chi, \psi)$, if its density is

$$f(x) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^{\lambda}}{2K_{\lambda}(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right), \quad x > 0,$$

where K_{λ} denotes a modified Bessel function of the third kind with index λ and the parameters satisfy $\chi > 0, \psi \geq 0$ if $\lambda < 0$; $\chi > 0, \psi > 0$ if $\lambda = 0$ and $\chi \geq 0, \psi > 0$ if $\lambda > 0$. For more on this Bessel function see [Abramowitz and Stegun, 1965].

The GIG density actually contains the gamma and inverse gamma densities as special limiting cases, corresponding to $\chi = 0$ and $\psi = 0$ respectively. Thus, when $\boldsymbol{\gamma} = \mathbf{0}$ and $\psi = 0$ the mixture distribution in (14) is multivariate t.

D3. Generalized Hyperbolic Distributions

The generalised hyperbolic density $f(\mathbf{x}) \propto$

$$\frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{(\chi + Q(\mathbf{x};\boldsymbol{\mu},\Sigma))(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)\exp\left((\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}\right)}{\left(\sqrt{(\chi + Q(\mathbf{x};\boldsymbol{\mu},\Sigma))(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}-\lambda}}.$$

where

$$Q(\mathbf{x};\boldsymbol{\mu},\Sigma) = (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$$

and the normalising constant is

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

Notes on Generalized Hyperbolic

- Notation: $\mathbf{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$.

- The class is closed under linear operations (including marginalization). If $\mathbf{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ and we consider $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$ where $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$ then $\mathbf{Y} \sim \mathrm{GH}_k(\lambda, \chi, \psi, B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B', B\boldsymbol{\gamma})$. A version of the variance-covariance method may be based on this family.

- The distribution may be fitted to data using the EM algorithm. Note that there is an identifiability problem (too many parameters) that is usually solved by setting $|\Sigma| = 1$. [McNeil et al., 2004]

Special Cases

- If $\lambda = 1$ we get a multivariate distribution whose univariate margins are one-dimensional hyperbolic distributions, a model widely used in univariate analyses of financial return data.

- If $\lambda = -1/2$ then the distribution is known as a normal inverse Gaussian (NIG) distribution. This model has also been used in univariate analyses of return data; it's functional form is similar to the hyperbolic with a slightly heavier tail.

- If $\lambda > 0$ and $\chi = 0$ we get a limiting case of the distribution known variously as a generalised Laplace, Bessel function or variance gamma distribution.

- If $\lambda = -\nu/2$, $\chi = \nu$ and $\psi = 0$ we get an asymmetric or skewed t distribution.

D4. Elliptical distributions

A random vector $(Y_1, \ldots, Y_d)$ is spherical if its distribution is invariant under rotations, i.e. for all $U \in \mathbb{R}^{d \times d}$ with $U'U = UU' = I_d$

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} U\mathbf{Y}.$$

A random vector $(X_1, \ldots, X_d)$ is called elliptical if it is an affine transform of a spherical random vector $(Y_1, \ldots, Y_k)$,

$$\mathbf{X} = A\mathbf{Y} + \mathbf{b},$$

$A \in \mathbb{R}^{d \times k}$, $\mathbf{b} \in \mathbb{R}^d$.

A normal variance mixture in (13) with $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I$ is spherical; any normal variance mixture is elliptical.

Properties of Elliptical Distributions

- The density of an elliptical distribution is constant on ellipsoids.
- Many of the nice properties of the multivariate normal are preserved. In particular, all linear combinations $a_1X_1 + \ldots + a_dX_d$ are of the same type.
- All marginal distributions are of the same type.
- Linear correlation matrices successfully summarise dependence, since mean vector, covariance matrix and the distribution type of the marginals determine the joint distribution uniquely.

Elliptical Distributions and Risk Management

Consider set of linear portfolios of elliptical risks
$\mathcal{P} = \{Z = \sum_{i=1}^{d} \lambda_i X_i \mid \sum_{i=1}^{d} \lambda_i = 1\}$.

- VaR is a coherent risk measure in this world. It is monotonic, positive homogeneous (P1), translation preserving (P2) and, most importantly, sub-additive
  $\mathrm{VaR}_\alpha(Z_1+Z_2) \leq \mathrm{VaR}_\alpha(Z_1)+\mathrm{VaR}_\alpha(Z_2)$, for $Z_1, Z_2 \in \mathcal{P}, \alpha > 0.5$.

- Among all portfolios with the same expected return, the portfolio minimizing VaR, or any other risk measure ϱ satisfying
  P1 $\varrho(\lambda Z) = \lambda \varrho(Z),\ \lambda \geq 0$,
  P2 $\varrho(Z + a) = \varrho(Z) + a,\ a \in \mathbb{R}$,
  is the Markowitz variance minimizing portfolio.
  Risk of portfolio takes the form $\varrho(Z) = E(Z) + \text{const} \cdot \mathrm{sd}(Z)$.

References

- [Barndorff-Nielsen and Shephard, 1998] (generalized hyperbolic distributions)
- [Barndorff-Nielsen, 1997] (NIG distribution)
- [Eberlein and Keller, 1995]) (hyperbolic distributions)
- [Prause, 1999] (GH distributions - PhD thesis)
- [Fang et al., 1987] (elliptical distributions)
- [Embrechts et al., 2001] (elliptical distributions in RM)

E. Copulas, Correlation and Extremal Dependence

1. Describing Dependence with Copulas
2. Survey of Useful Copula Families
3. Simulation of Copulas
4. Understanding the Limitations of Correlation
5. Tail dependence and other Alternative Dependence Measures
6. Fitting Copulas to Data

E1. Modelling Dependence with Copulas

On Uniform Distributions

Lemma 1: probability transform

Let X be a random variable with continuous distribution function F. Then $F(X) \sim U(0,1)$ (standard uniform).

$$P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F(F^{-1}(u)) = u, \quad \forall u \in (0,1).$$

Lemma 2: quantile transform

Let U be uniform and F the distribution function of any rv X. Then $F^{-1}(U) \overset{\text{d}}{=} X$ so that $P(F^{-1}(U) \leq x) = F(x)$.

These facts are the key to all statistical simulation and essential in dealing with copulas.

A Definition

A copula is a multivariate distribution function $C : [0,1]^d \to [0,1]$ with standard uniform margins (or a distribution with such a df).

Properties

- Uniform Margins
  $C(1,\ldots,1,u_i,1,\ldots,1) = u_i$ for all $i \in \{1,\ldots,d\}$, $u_i \in [0,1]$
- Fréchet Bounds

$$\max\left\{\sum_{i=1}^{d} u_i + 1 - d, 0\right\} \leq C(\mathbf{u}) \leq \min\{u_1,\ldots,u_d\}.$$

Remark: right hand side is df of $\overbrace{(U,\ldots,U)}^{\text{d times}}$, where $U \sim U(0,1)$.

Sklar's Theorem

Let F be a joint distribution function with margins $F_1, \ldots, F_d$. There exists a copula such that for all $x_1, \ldots, x_d$ in $[-\infty, \infty]$

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d)).$$

If the margins are continuous then C is unique; otherwise C is uniquely determined on Ran$F_1 \times$ Ran$F_2 \ldots \times$ RanF_d.

And conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then F defined above is a multivariate df with margins $F_1, \ldots, F_d$.

Idea of Proof in Continuous Case

Henceforth, unless explicitly stated, vectors $\mathbf{X}$ will be assumed to have continuous marginal distributions. In this case:

$$\begin{aligned} F(x_1,\ldots,x_d) &= P\left(X_1 \le x_1,\ldots,X_d \le x_d\right) \\ &= P\left(F_1(X_1) \le F_1(x_1),\ldots,F_d(X_d) \le F_d(x_d)\right) \\ &= C(F_1(x_1),\ldots,F_d(x_d)). \end{aligned}$$

The unique copula C can be calculated from $F, F_1, \ldots, F_d$ using

$$C(u_1,\ldots,u_d) = F\left(F_1^{-1}(u_1),\ldots,F_d^{-1}(u_d)\right).$$

Copulas and Dependence Structures

Sklar's theorem shows how a unique copula C fully describes the dependence of $\mathbf{X}$. This motivates a further definition.

Definition: Copula of $\mathbf{X}$

The copula of $(X_1, \ldots, X_d)$ (or F) is the df C of $(F_1(X_1), \ldots, F_d(X_d))$.

We sometimes refer to C as the dependence structure of F.

Invariance

C is invariant under strictly increasing transformations of the marginals.

If $T_1, \ldots, T_d$ are strictly increasing, then $(T_1(X_1), \ldots, T_d(X_d))$ has the same copula as $(X_1, \ldots, X_d)$.

Examples of copulas

- Independence

  $X_1, \ldots, X_d$ are mutually independent $\iff$ their copula C satisfies $C(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i$.

- Comonotonicity - perfect dependence

  $X_i \overset{\text{a.s.}}{=} T_i(X_1)$, T_i strictly increasing, $i = 2, \ldots, d$, $\iff$ C satisfies

  $C(u_1, \ldots, u_d) = \min\{u_1, \ldots, u_d\}$.

- Countermonotonicity - perfect negative dependence (d=2)

  $X_2 \overset{\text{a.s.}}{=} T(X_1)$, T strictly decreasing, $\iff$ C satisfies $C(u_1, u_2) = \max\{u_1 + u_2 - 1, 0\}$.

Parametric Copulas

There are basically two possibilities:

- Copulas implicit in well-known parametric distributions
  Recall $C(u_1, \ldots, u_d) = F\left(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right)$.
- Closed-form parametric copula families.

Gaussian Copula: an implicit copula

Let $\mathbf{X}$ be standard multivariate normal with correlation matrix P.

$$\begin{aligned} C_P^{\text{Ga}}(u_1, \ldots, u_d) &= P\left(\Phi(X_1) \leq u_1, \ldots, \Phi(X_d) \leq u_d\right) \\ &= P\left(X_1 \leq \Phi^{-1}(u_1), \ldots, X_d \leq \Phi^{-1}(u_d)\right) \end{aligned}$$

where Φ is df of standard normal.

$P = I$ gives independence; as $P \to J$ we get comonotonicity.

E2. Parametric Copula Families

Elliptical or Normal Mixture Copulas

The Gaussian copula is an elliptical copula. Using a similar approach we can extract copulas from other multivariate normal mixture distributions.

Examples

- The t copula $C^{\mathrm{t}}_{\nu,P}$
- The generalised hyperbolic copula

The elliptical copulas are rich in parameters - parameter for every pair of variables; easy to simulate.

Archimedean Copulas $d = 2$

These have simple closed forms and are useful for calculations. However, higher dimensional extensions are not rich in parameters.

- Gumbel Copula

$$C_\beta^{Gu}(u_1, u_2) = \exp\left(-\left((-\log u_1)^\beta + (-\log u_2)^\beta\right)^{1/\beta}\right).$$

$\beta \geq 1$: $\beta = 1$ gives independence; $\beta \to \infty$ gives comonotonicity.

- Clayton Copula

$$C_\beta^{Cl}(u_1, u_2) = \left(u_1^{-\beta} + u_2^{-\beta} - 1\right)^{-1/\beta}.$$

$\beta > 0$: $\beta \to 0$ gives independence ; $\beta \to \infty$ gives comonotonicity.

Archimedean Copulas in Higher Dimensions

All our Archimedean copulas have the form

$$C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2)),$$

where $\psi : [0, 1] \mapsto [0, \infty]$ is strictly decreasing and convex with $\psi(1) = 0$ and $\lim_{t \to 0} \psi(t) = \infty$.
The simplest higher dimensional extension is

$$C(u_1, \ldots, u_d) = \psi^{-1}(\psi(u_1) + \cdots + \psi(u_d)).$$

Example: Gumbel copula: $\psi(t) = -(\log(t))^\beta$

$$C_\beta^{\mathsf{Gu}}(u_1, \ldots, u_d) = \exp\left(- \left((-\log u_1)^\beta + \cdots + (-\log u_d)^\beta \right)^{1/\beta} \right).$$

These copulas are exchangeable (invariant under permutations).

E3. Simulating Copulas

Normal Mixture (Elliptical) Copulas

Simulating Gaussian copula C_P^{Ga}

- Simulate $\mathbf{X} \sim N_d(\mathbf{0}, P)$
- Set $\mathbf{U} = (\Phi(X_1), \ldots, \Phi(X_d))'$ (probability transformation)

Simulating t copula $C_{\nu,P}^{\text{t}}$

- Simulate $\mathbf{X} \sim t_d(\nu, \mathbf{0}, P)$
- $\mathbf{U} = (t_\nu(X_1), \ldots, t_\nu(X_d))'$ (probability transformation)
  t_ν is df of univariate t distribution.

Meta–Gaussian and Meta–t Distributions

If $(U_1, \ldots, U_d) \sim C_P^{\text{Ga}}$ and F_i are univariate dfs other than univariate normal then

$$\left(F_1^{\leftarrow}(U_1), \ldots, F_d^{\leftarrow}(U_d)\right)$$

has a meta–Gaussian distribution. Thus it is easy to simulate vectors with the Gaussian copula and arbitrary margins.

In a similar way we can construct and simulate from meta t_ν distributions. These are distributions with copula $C_{\nu,P}^{\text{t}}$ and margins other than univariate t_ν.

Simulating Archimedean Copulas

For the most useful of the Archimedean copulas (such as Clayton and Gumbel) techniques exist to simulate the exchangeable versions in arbitrary dimensions. The theory on which this is based may be found in Marshall and Olkin (1988).

Algorithm for d-dimensional Clayton copula C_β^{Cl}

- Simulate a gamma variate X with parameter $\alpha = 1/\beta$. This has density $f(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$.
- Simulate d independent standard uniforms $U_1, \ldots, U_d$.
- Return $\left(\left(1 - \frac{\log U_1}{X}\right)^{-1/\beta}, \ldots, \left(1 - \frac{\log U_d}{X}\right)^{-1/\beta}\right)$.

E4. Understanding Limitations of Correlation

Drawbacks of Linear Correlation

Denote the linear correlation of two random variables X_1 and X_2 by $\rho(X_1, X_2)$. We should be aware of the following.

- Linear correlation only gives a scalar summary of (linear) dependence and $\mathrm{var}(X_1), \mathrm{var}(X_2)$ must exist.
- X_1, X_2 independent $\Rightarrow \rho(X, Y) = 0$.
  But $\rho(X_1, X_2) = 0 \not\Rightarrow X_1, X_2$ independent.
  Example: spherical bivariate t-distribution with ν d.f.
- Linear correlation is not invariant with respect to strictly increasing transformations T of X_1, X_2, i.e. generally

$$\rho(T(X_1), T(X_2)) \neq \rho(X_1, X_2).$$

A Fallacy in the Use of Correlation

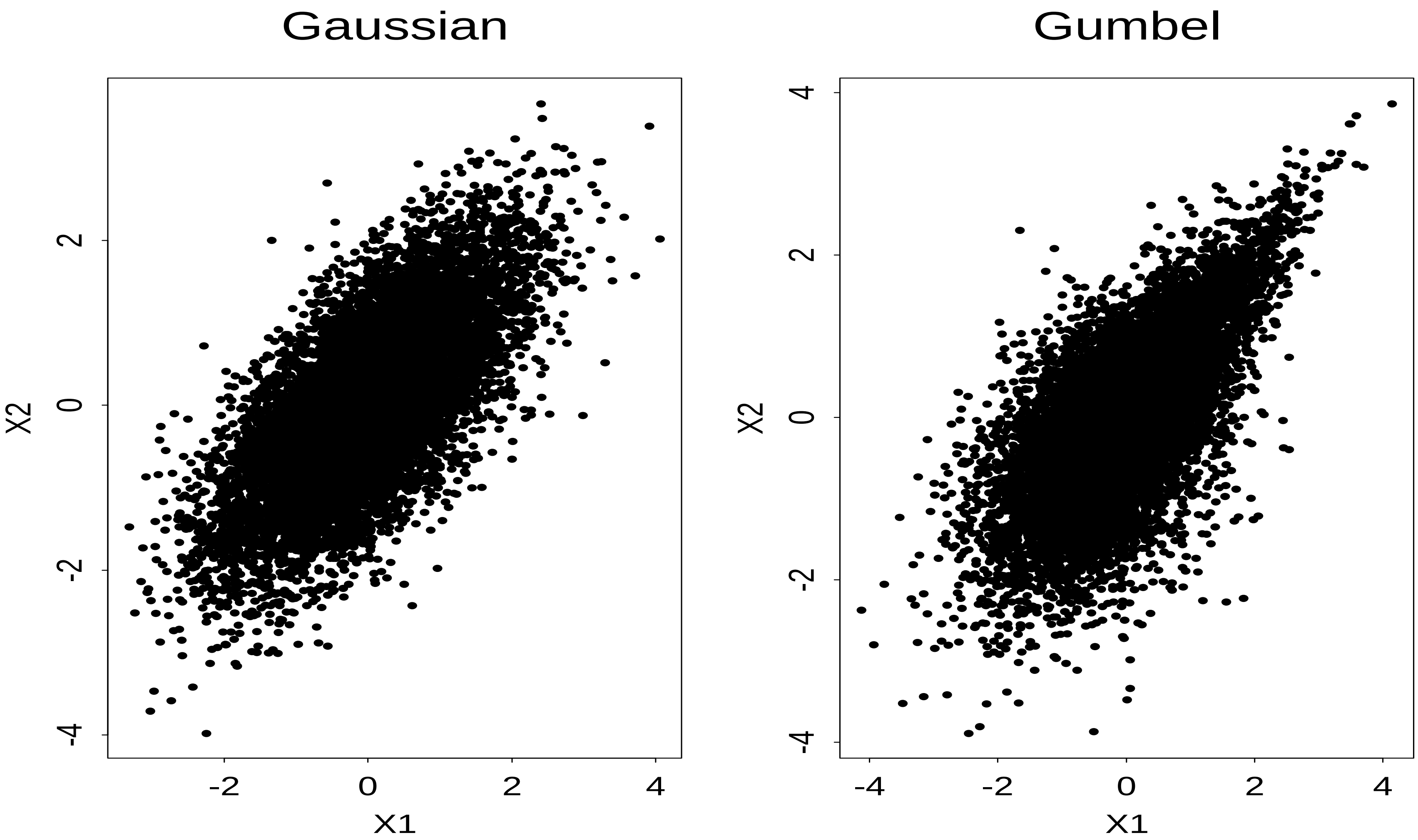
Consider the random vector $(X_1, X_2)'$.

"Marginal distributions and correlation determine the joint distribution".

- True for the class bivariate normal distributions or, more generally, for elliptical distributions.

- Wrong in general, as the next example shows.

Gaussian and Gumbel Copulas Compared



Margins are standard normal; correlation is 70%.

E5. Alternative Dependence Concepts

Rank Correlation (let C denote copula of X_1 and X_2)

Spearman's rho

$$\begin{aligned} \rho_S(X_1, X_2) &= \rho(F_1(X_1), F_2(X_2)) = \rho(\text{copula}) \\ \rho_S(X_1, X_2) &= 12 \int_0^1 \int_0^1 \{C(u_1, u_2) - u_1 u_2\} du_1 du_2. \end{aligned}$$

Kendall's tau

Take an independent copy of (X_1, X_2) denoted $(\widetilde{X}_1, \widetilde{X}_2)$.

$$\begin{aligned} \rho_\tau(X_1, X_2) &= 2P\left((X_1 - \widetilde{X}_1)(X_2 - \widetilde{X}_2) > 0\right) - 1 \\ \rho_\tau(X_1, X_2) &= 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1. \end{aligned}$$

Properties of Rank Correlation

(not shared by linear correlation)

True for Spearman's rho (ρ_S) or Kendall's tau (ρ_τ).

- ρ_S depends only on copula of $(X_1, X_2)'$.
- ρ_S is invariant under strictly increasing transformations of the random variables.
- $\rho_S(X_1, X_2) = 1 \iff X_1, X_2$ comonotonic.
- $\rho_S(X_1, X_2) = -1 \iff X_1, X_2$ countermonotonic.

Kendall's Tau in Elliptical Models

Suppose $\mathbf{X} = (X_1, X_2)'$ has any elliptical distribution; for example $\mathbf{X} \sim t_2(\nu, \boldsymbol{\mu}, \Sigma)$. Then

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin\left(\rho(X_1, X_2)\right). \qquad (17)$$

Remarks:

1. In case of infinite variances we simply interpret $\rho(X_1, X_2)$ as $\Sigma_{1,2}/\sqrt{\Sigma_{1,1}\Sigma_{2,2}}$.
2. Result of course implies that if $\mathbf{Y}$ has copula $C^t_{\nu,P}$ then $\rho_\tau(Y_1, Y_2) = \frac{2}{\pi}\arcsin(P_{1,2})$.
3. An estimator of ρ_τ is given by

$$\widehat{\rho_\tau}(X_1, X_2) = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \operatorname{sgn}\left[\left(X_{i,1} - X_{j,1}\right)\left(X_{i,2} - X_{j,2}\right)\right].$$

Tail Dependence or Extremal Dependence

Objective: measure dependence in joint tail of bivariate distribution. When limit exists, coefficient of upper tail dependence is

$$\lambda_u(X_1, X_2) = \lim_{q \to 1} P(X_2 > \mathrm{VaR}_q(X_2) \mid X_1 > \mathrm{VaR}_q(X_1)),$$

Analogously the coefficient of lower tail dependence is

$$\lambda_\ell(X_1, X_2) = \lim_{q \to 0} P\left(X_2 \leq \mathrm{VaR}_q(X_2) \mid X_1 \leq \mathrm{VaR}_q(X_1)\right).$$

These are functions of the copula given by

$$\begin{aligned} \lambda_u &= \lim_{q \to 1} \frac{\overline{C}(q,q)}{1-q} = \lim_{q \to 1} \frac{1 - 2q + C(q,q)}{1-q}, \\ \lambda_\ell &= \lim_{q \to 0} \frac{C(q,q)}{q}. \end{aligned}$$

Tail Dependence

Clearly $\lambda_u \in [0,1]$ and $\lambda_\ell \in [0,1]$.

For elliptical copulas $\lambda_u = \lambda_\ell =: \lambda$. True of all copulas with radial symmetry: $(U_1, U_2) \stackrel{\mathrm{d}}{=} (1 - U_1, 1 - U_2)$.

Terminology:

$\lambda_u \in (0,1]$: upper tail dependence,

$\lambda_u = 0$: asymptotic independence in upper tail,

$\lambda_\ell \in (0,1]$: lower tail dependence,

$\lambda_\ell = 0$: asymptotic independence in lower tail.

Examples of tail dependence

The Gaussian copula is asymptotically independent for $|\rho| < 1$.

The t copula is tail dependent when $\rho > -1$.

$$\lambda = 2\bar{t}_{\nu+1}\left(\sqrt{\nu+1}\sqrt{1-\rho}/\sqrt{1+\rho}\right).$$

The Gumbel copula is upper tail dependent for $\beta > 1$.

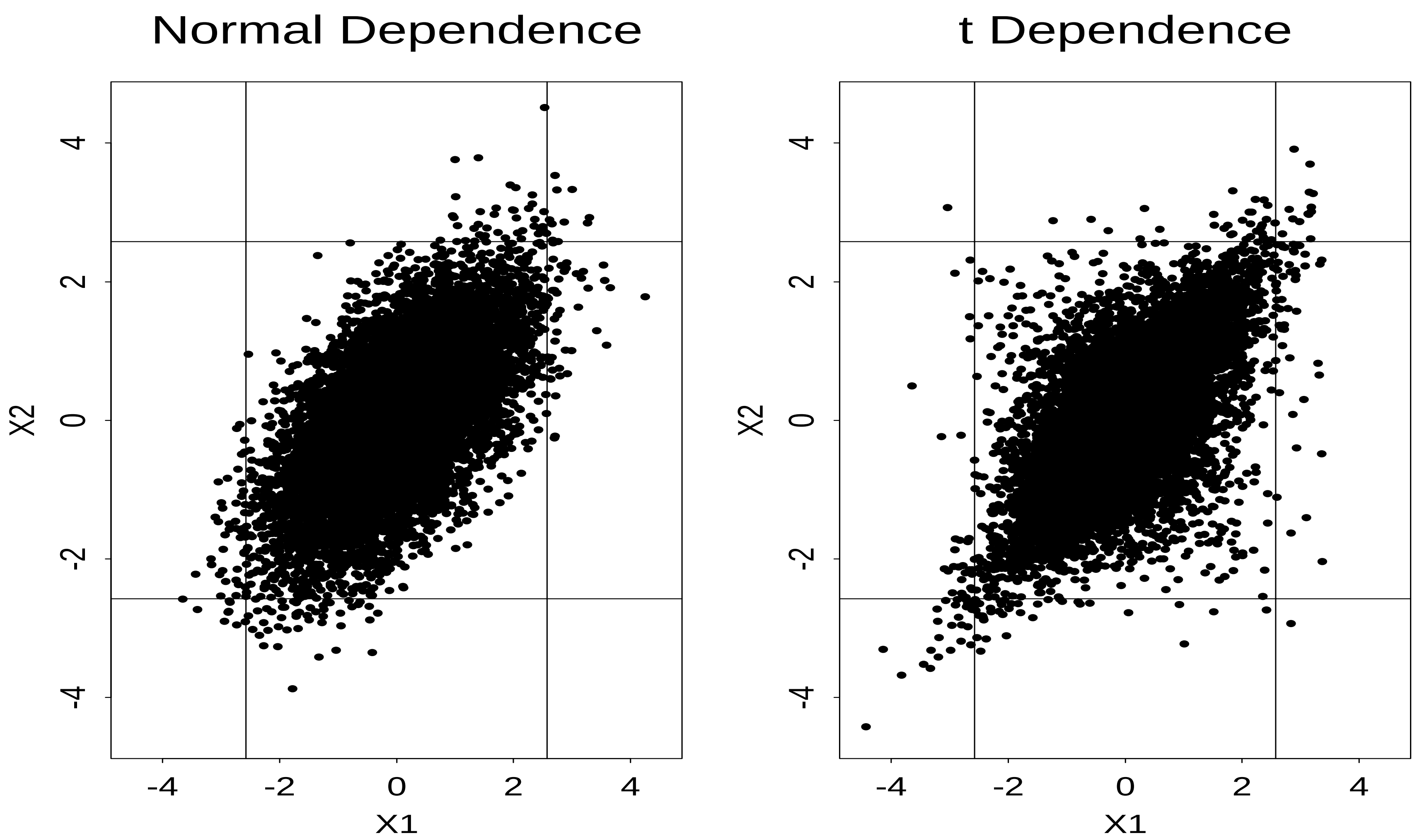
$$\lambda_u = 2 - 2^{1/\beta}.$$

The Clayton copula is lower tail dependent for $\beta > 0$.

$$\lambda_\ell = 2^{-1/\beta}.$$

Recall dependence model in Fallacy 1b: $\lambda_u = \lambda_\ell = 0.5$.

Gaussian and t3 Copulas Compared



Copula parameter $\rho = 0.7$; quantiles lines 0.5% and 99.5%.

Joint Tail Probabilities at Finite Levels

ρ	C	Quantile			
		95%	99%	99.5%	99.9%
0.5	N	1.21×10^{-2}	1.29×10^{-3}	4.96×10^{-4}	5.42×10^{-5}
0.5	t8	1.20	1.65	1.94	3.01
0.5	t4	1.39	2.22	2.79	4.86
0.5	t3	1.50	2.55	3.26	5.83
0.7	N	1.95×10^{-2}	2.67×10^{-3}	1.14×10^{-3}	1.60×10^{-4}
0.7	t8	1.11	1.33	1.46	1.86
0.7	t4	1.21	1.60	1.82	2.52
0.7	t3	1.27	1.74	2.01	2.83

For normal copula probability is given.
For t copulas the factor by which Gaussian probability must be multiplied is given.

Joint Tail Probabilities, $d \geq 2$

ρ	C	Dimension d			
		2	3	4	5
0.5	N	1.29×10^{-3}	3.66×10^{-4}	1.49×10^{-4}	7.48×10^{-5}
0.5	t8	1.65	2.36	3.09	3.82
0.5	t4	2.22	3.82	5.66	7.68
0.5	t3	2.55	4.72	7.35	10.34
0.7	N	2.67×10^{-3}	1.28×10^{-3}	7.77×10^{-4}	5.35×10^{-4}
0.7	t8	1.33	1.58	1.78	1.95
0.7	t4	1.60	2.10	2.53	2.91
0.7	t3	1.74	2.39	2.97	3.45

We consider only 99% quantile and case of equal correlations.

Financial Interpretation

Consider daily returns on five financial instruments and suppose that we believe that all correlations between returns are equal to 50%. However, we are unsure about the best multivariate model for these data.

If returns follow a multivariate Gaussian distribution then the probability that on any day all returns fall below their 1% quantiles is 7.48×10^{-5}. In the long run such an event will happen once every 13369 trading days on average, that is roughly once every 51.4 years (assuming 260 trading days in a year).

On the other hand, if returns follow a multivariate t distribution with four degrees of freedom then such an event will happen 7.68 times more often, that is roughly once every 6.7 years.

E6. Fitting Copulas to Data

Situation

We have identically distributed data vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from a distribution with unknown (continuous) margins $F_1, \ldots, F_d$ and with unknown copula C. We adopt a two-stage estimation procedure.

Stage 1

Estimate marginal distributions either with

1. parametric models $\widehat{F}_1, \ldots, \widehat{F}_d$,

2. a form of the empirical distribution function such as $\widehat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} 1_{\{X_{i,j} \leq x\}}, \quad j = 1, \ldots, d$,

3. empirical df with EVT tail model.

Stage 2: Estimating the Copula

We form a pseudo-sample of observations from the copula

$$\widehat{\mathbf{U}}_i = \left(\widehat{U}_{i,1}, \ldots, \widehat{U}_{i,d}\right)' = \left(\widehat{F}_1(X_{i,1}), \ldots, \widehat{F}_d(X_{i,d})\right)', \quad i = 1, \ldots, n.$$

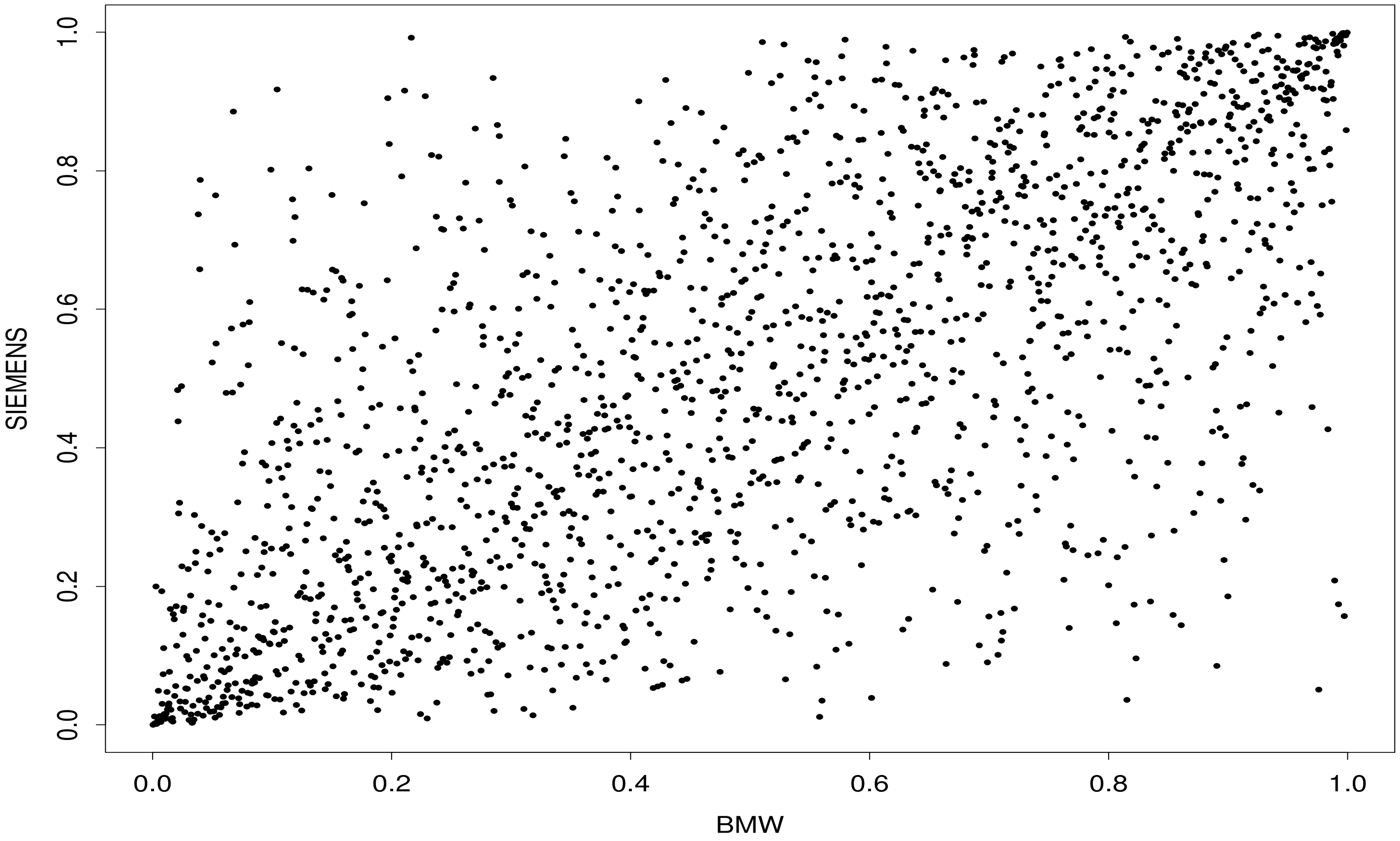
and fit parametric copula C by maximum likelihood.

Copula density is $c(u_1, \ldots, u_d; \boldsymbol{\theta}) = \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_d} C(u_1, \ldots, u_d; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denote unknown parameters. The log-likelihood is

$$l(\boldsymbol{\theta}; \widehat{\mathbf{U}}_1, \ldots, \widehat{\mathbf{U}}_n) = \sum_{i=1}^{n} \log c(\widehat{U}_{i,1}, \ldots, \widehat{U}_{i,d}; \boldsymbol{\theta}).$$

Independence of vector observations assumed for simplicity. More theory is found in Genest and Rivest (1993) and Maschal and Zeevi (2002).

BMW-Siemens Example: Stage 1



The pseudo-sample from copula after estimation of margins.

Stage 2: Parametric Fitting of Copulas

Copula	ρ, β	ν	std.error(s)	log-likelihood
Gauss	0.70		0.0098	610.39
t	0.70	4.89	0.0122,0.73	649.25
Gumbel	1.90		0.0363	584.46
Clayton	1.42		0.0541	527.46

Goodness-of-fit.
Akaike's criterion (AIC) suggests choosing model that minimises

$$\text{AIC} = 2p - 2 \cdot (\text{log-likelihood}),$$

where $p =$ number of parameters of model. This is clearly t model.

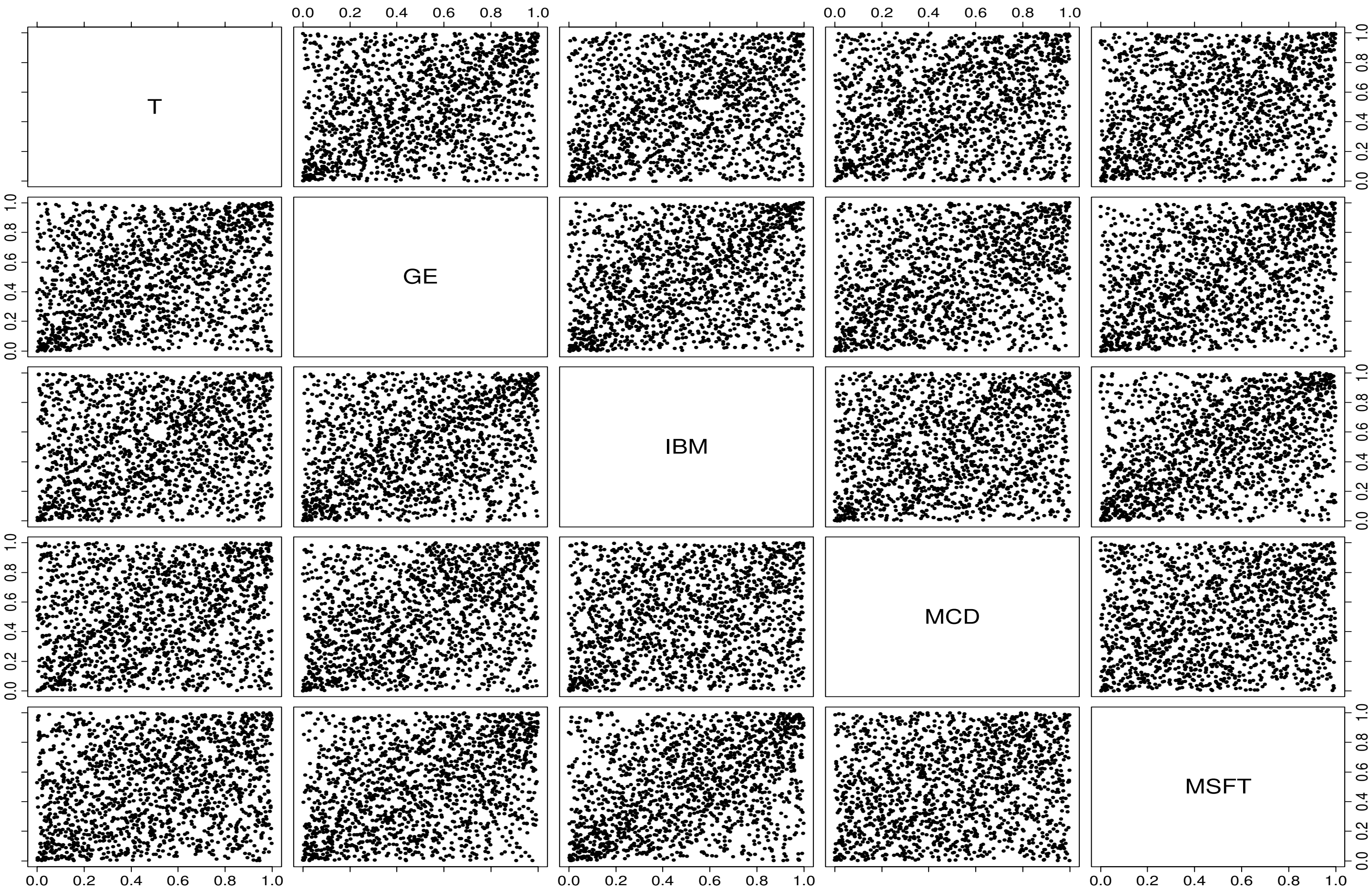
Remark. Formal methods for goodness-of-fit also available.

Fitting the t or Gaussian Copulas

ML estimation may be difficult in very high dimensions, due to the large number of parameters these copulas possess. As an alternative we can use the rank correlation calibration methods described earlier. For the t copula a hybrid method is possible:

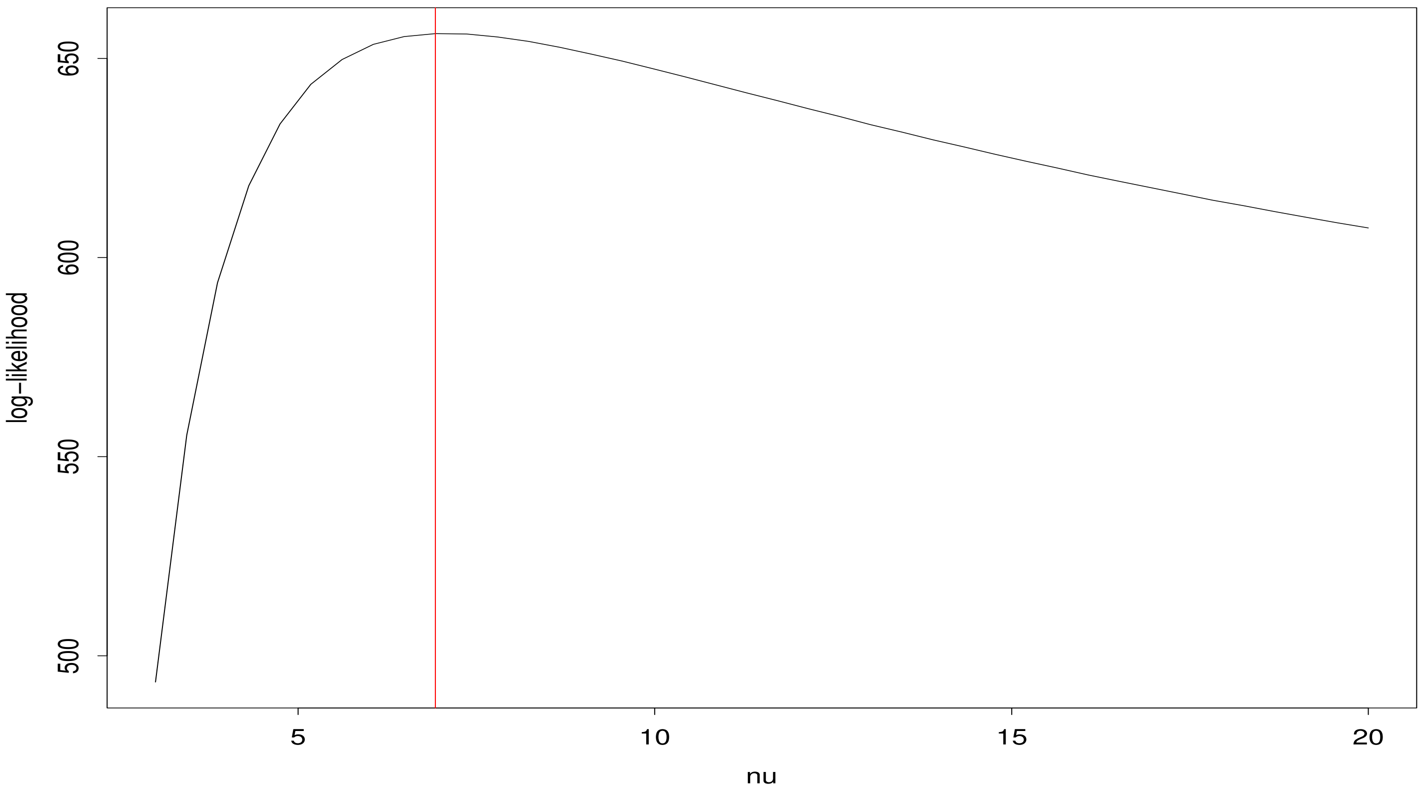
- Estimate Kendall's tau matrix from the data.
- Recall that if $\mathbf{X}$ is meta-t with df $C^t_{\nu,P}(F_1,\ldots,F_d)$ then $\rho_\tau(X_i,X_j)=\frac{2}{\pi}\arcsin(P_{i,j})$. Follows from (17).
- Estimate $\widehat{P}_{i,j}=\sin\left(\frac{\pi}{2}\widehat{\rho}_\tau(X_i,X_j)\right)$. Check positive definiteness!
- Estimate remaining parameter ν by the ML method.

Dow Jones Example: Stage 1



The pseudo-sample from copula after estimation of margins.

Stage 2: Fitting the t Copula



Daily returns on ATT, General Electric, IBM, McDonalds, Microsoft.
Form of likelihood for nu indicates non-Gaussian dependence.

References

- [Embrechts et al., 2001] (dependence and copulas in RM)
- [Joe, 1997] (on dependence in general)
- [Nelsen, 1999] (standard reference on bivariate copulas)
- [Lindskog, 2000] (useful supplementary reading)
- [Marshall and Olkin, 1988] (simulation of Archimedean copulas)
- [Klugman and Parsa.R., 1999] (copula fitting in insurance)
- [Frees and Valdez, 1997] (role of copulas in insurance)
- [Genest and Rivest, 1993] (theory of copula fitting)
- [Mashal and Zeevi, 2002] (copula fitting in finance)

Bibliography

[Abramowitz and Stegun, 1965] Abramowitz, M. and Stegun, I., editors (1965). *Handbook of Mathematical Functions*. Dover Publications, Inc., New York.

[Artzner et al., 1999] Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999). Coherent measures of risk. *Math. Finance*, 9:203–228.

[Atkinson, 1982] Atkinson, A. (1982). The simulation of generalized inverse gaussian and hyperbolic random variables. *SIAM J. Sci. Comput.*, 3(4):502–515.

[Barndorff-Nielsen, 1997] Barndorff-Nielsen, O. (1997). Normal inverse gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.*, 24:1–13.

[Barndorff-Nielsen and Shephard, 1998] Barndorff-Nielsen, O. and Shephard, N. (1998). Aggregation and model construction for volatility models. Preprint, Center for Analytical Finance, University of Aarhus.

[Crouhy et al., 2001] Crouhy, M., Galai, D., and Mark, R. (2001). *Risk Management*. McGraw-Hill, New York.

[Eberlein and Keller, 1995] Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, 1:281–299.

[Eberlein et al., 1998] Eberlein, E., Keller, U., and Prause, K. (1998). New insights into smile, mispricing, and value at risk: the hyperbolic model. *J. Bus.*, 38:371–405.

[Embrechts et al., 2001] Embrechts, P., McNeil, A., and

Straumann, D. (2001). Correlation and dependency in risk management: properties and pitfalls. In Dempster, M. and Moffatt, H., editors, *Risk Management: Value at Risk and Beyond*, pages 176–223. Cambridge University Press, http://www.math.ethz.ch/~mcneil.

[Fang et al., 1987] Fang, K.-T., Kotz, S., and Ng, K.-W. (1987). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London.

[Frees and Valdez, 1997] Frees, E. and Valdez, E. (1997). Understanding relationships using copulas. *N. Amer. Actuarial J.*, 2(1):1–25.

[Genest and Rivest, 1993] Genest, C. and Rivest, L. (1993).

Statistical inference procedures for bivariate archimedean copulas. *J. Amer. Statist. Assoc.*, 88:1034–1043.

[Joe, 1997] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London.

[Jorion, 2001] Jorion, P. (2001). *Value at Risk: the New Benchmark for Measuring Financial Risk*. McGraw-Hill, New York, 2nd edition edition.

[Klugman and Parsa.R., 1999] Klugman, S. and Parsa.R. (1999). Fitting bivariate loss distributions with copulas. *Ins.: Mathematics Econ.*, 24:139–148.

[Kotz et al., 2000] Kotz, S., Balakrishnan, N., and Johnson, N. (2000). *Continuous Multivariate Distributions*. Wiley, New York.

[Lindskog, 2000] Lindskog, F. (2000). Modelling dependence with copulas. RiskLab Report, ETH Zurich.

[Mardia et al., 1979] Mardia, K., Kent, J., and Bibby, J. (1979). *Multivariate Analysis*. Academic Press, London.

[Marshall and Olkin, 1988] Marshall, A. and Olkin, I. (1988). Families of multivariate distributions. *J. Amer. Statist. Assoc.*, 83:834–841.

[Mashal and Zeevi, 2002] Mashal, R. and Zeevi, A. (2002). Beyond correlation: extreme co–movements between financial assets. Unpublished, Columbia University.

[McNeil et al., 2004] McNeil, A., Frey, R., and Embrechts, P. (2004). *Quantitative Risk Management: Concepts, Techniques and Tools*. `www.math.ethz.ch/∼mcneil/book.html`.

[Nelsen, 1999] Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer, New York.

[Prause, 1999] Prause, K. (1999). *The generalized hyperbolic model: estimation, financial derivatives and risk measures*. PhD thesis, Institut für Mathematische Statistik, Albert–Ludwigs–Universität Freiburg.

[Seber, 1984] Seber, G. (1984). *Multivariate Observations*. Wiley, New York.