Quantitative Risk Management

Concepts, Techniques and Tools

A.J. McNeil, R. Frey and P. Embrechts

Princeton University Press, 2005

Paul Embrechts

with the assistance of Valerie Chavez-Demoulin and Johanna Neslehova

`www.math.ethz.ch/~embrechts`

Lectures at the Federal Reserve Bank of Boston
September 26-30, 2005

Contents

H. Copulas and Dependence

I. EVT I: Maxima and Worst Case Losses

J. EVT II: Modelling Threshold Exceedances

K. EVT III: Advanced Topics

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

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

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

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

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

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

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

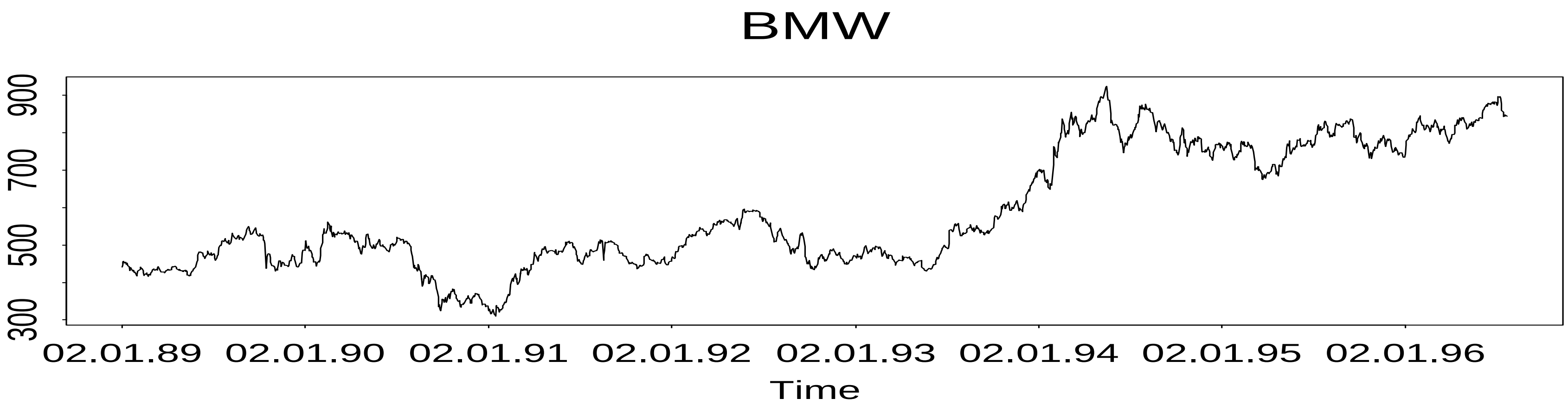
Consider d stocks; let α_i denote number of shares in stock i at time t and let $S_{t,i}$ denote price.

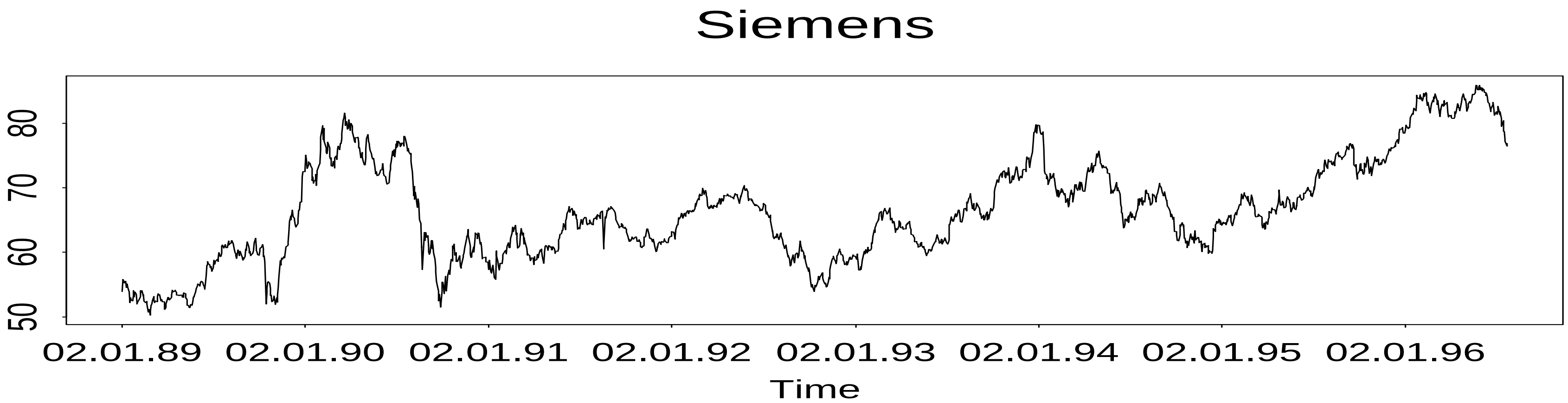
The risk factors: following standard convention we take logarithmic prices as risk factors $Z_{t,i} = \log S_{t,i}, 1 \leq i \leq d$.

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)

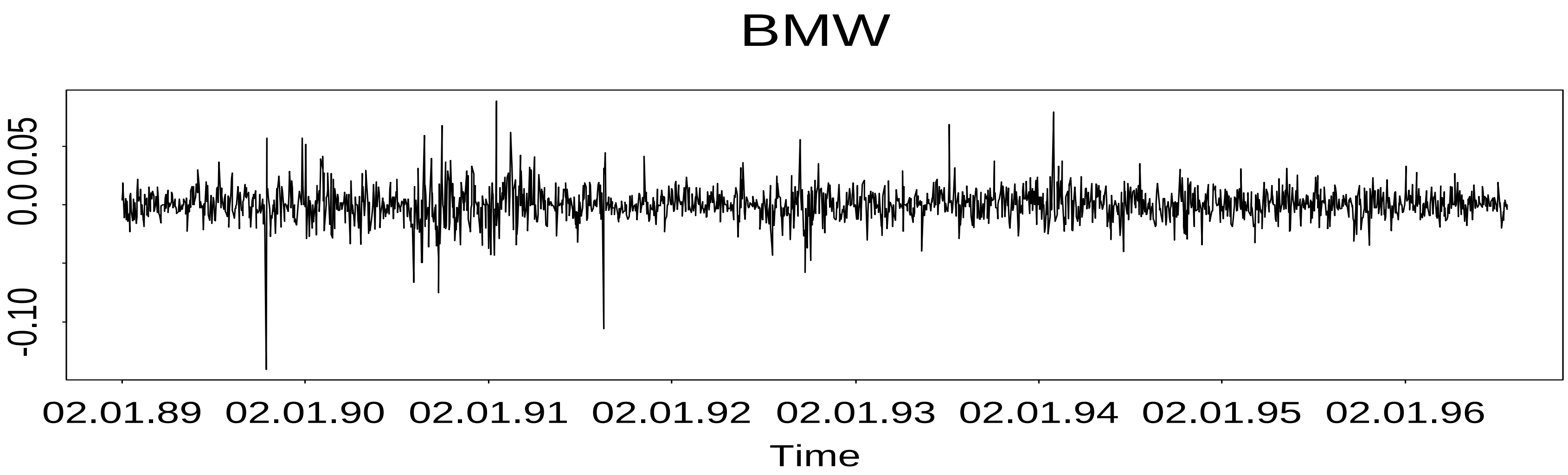
$$V_t = \sum_{i=1}^{d} \alpha_i S_{t,i} = \sum_{i=1}^{d} \alpha_i e^{Z_{t,i}}. \tag{4}$$

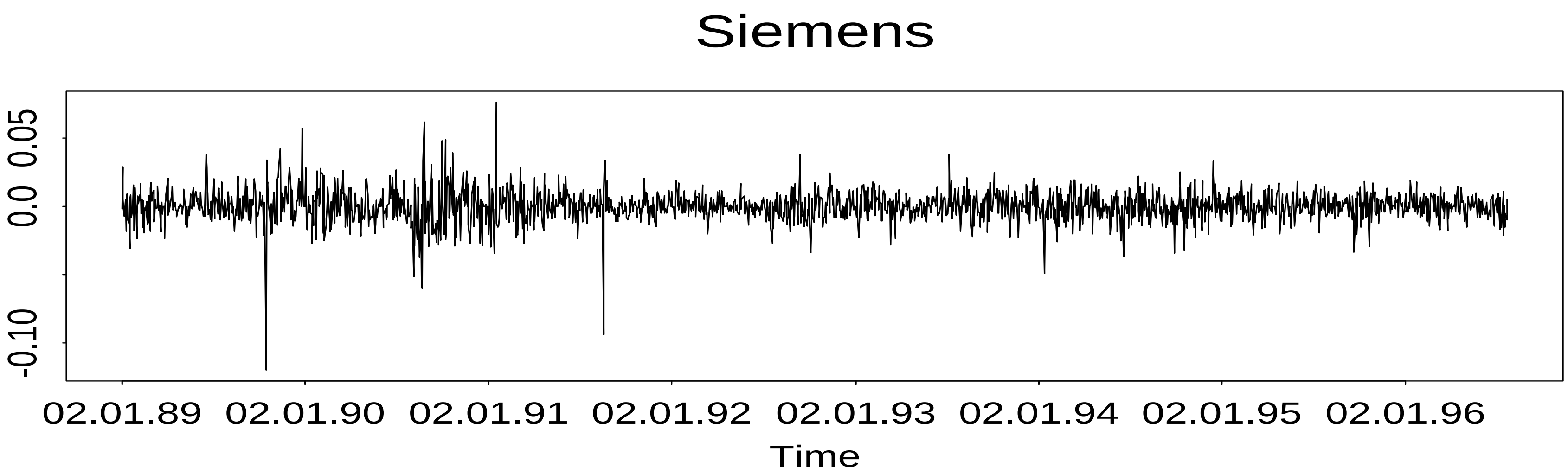




BMW and Siemens Data: 1972 days to 23.07.96.

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)

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

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

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

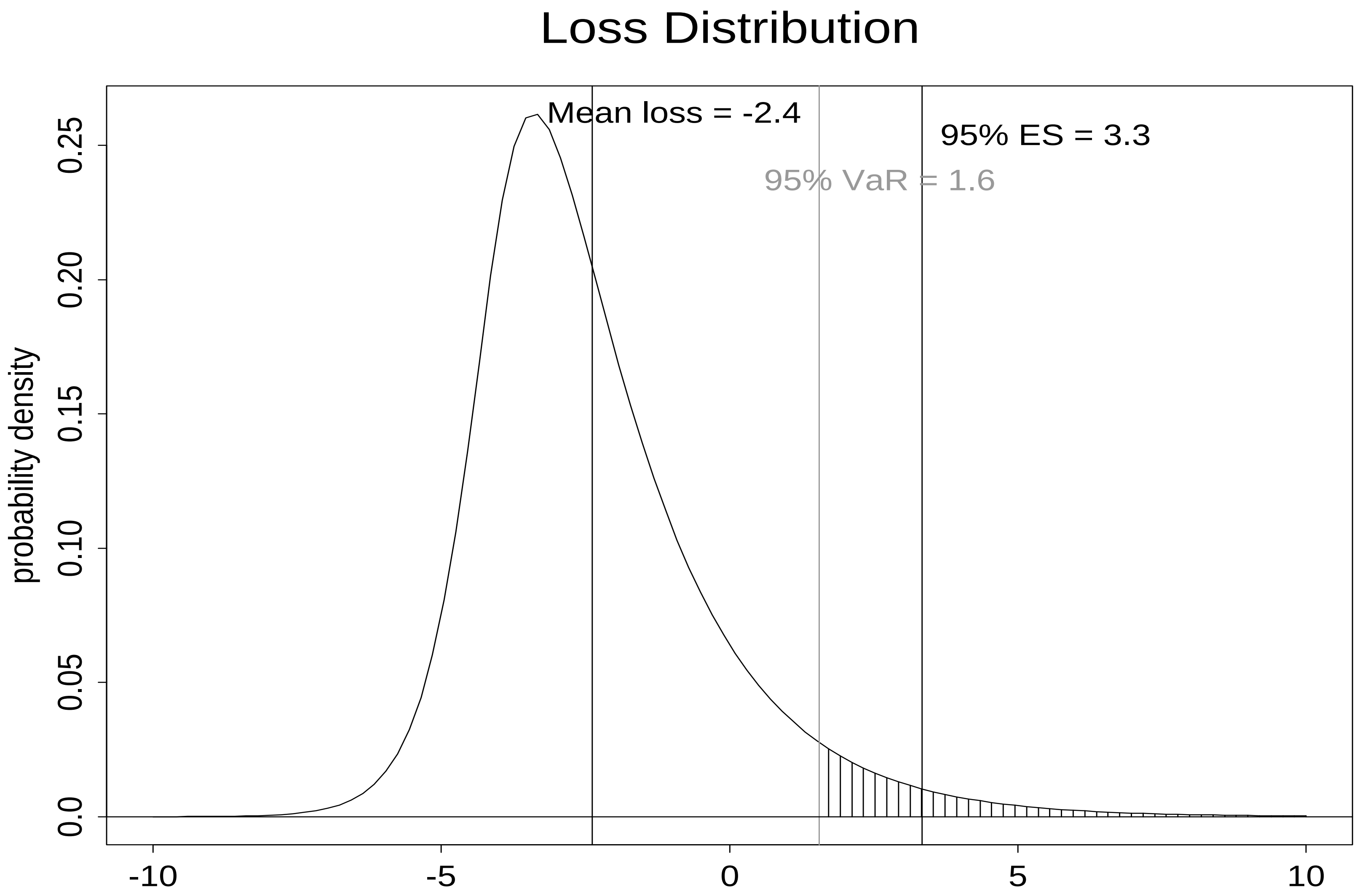
i.e. the α-quantile of F_L.

- Alternative risk measure: Expected shortfall defined as

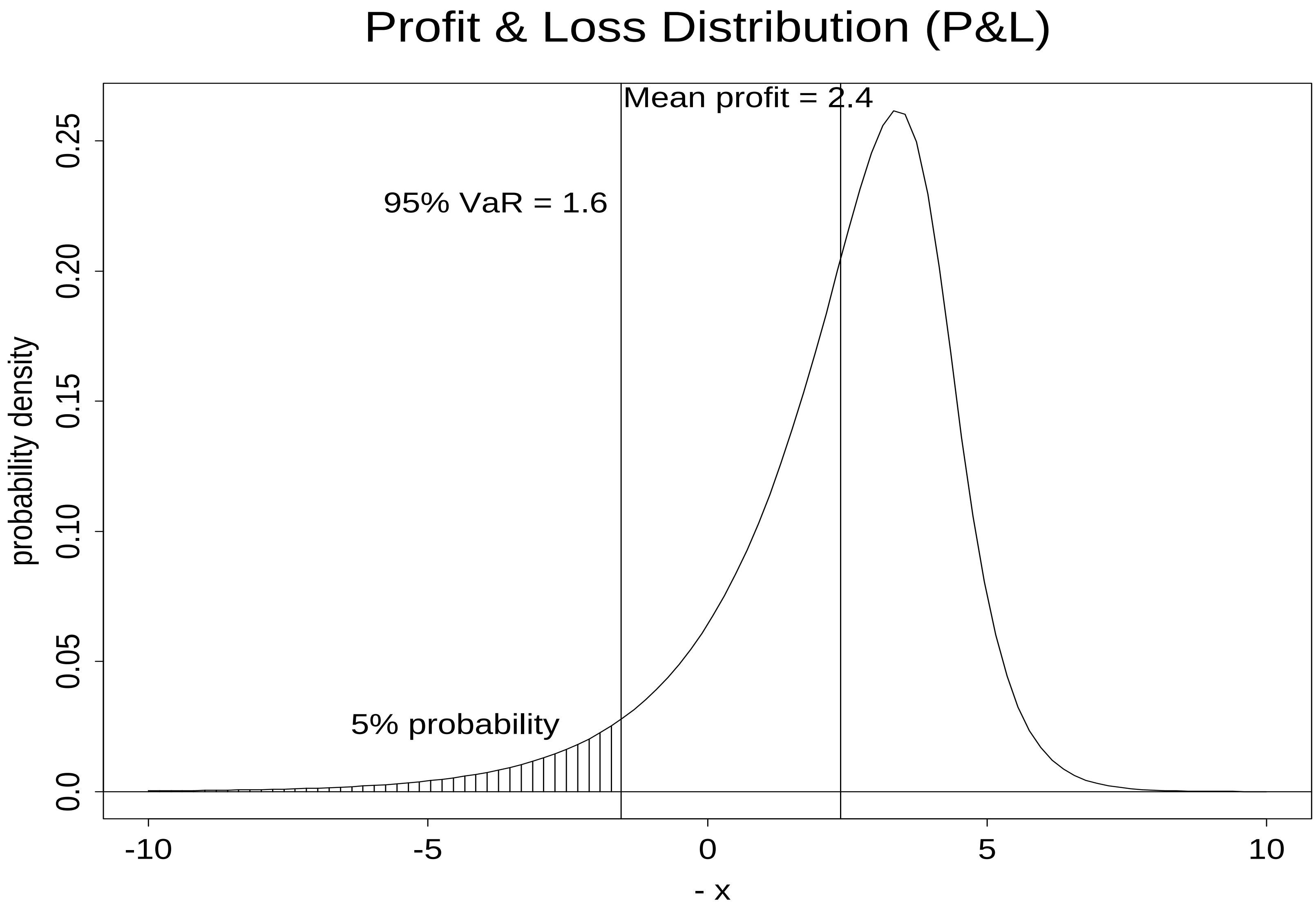
$$\mathrm{ES}_\alpha = E\left(L \;\middle|\; L > \mathrm{VaR}_\alpha\right)\,, \tag{7}$$

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

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

- ⋆ 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

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

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

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

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

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

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

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.

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 $\mathbf{X}_{t-n+1}, \ldots, \mathbf{X}_t$.

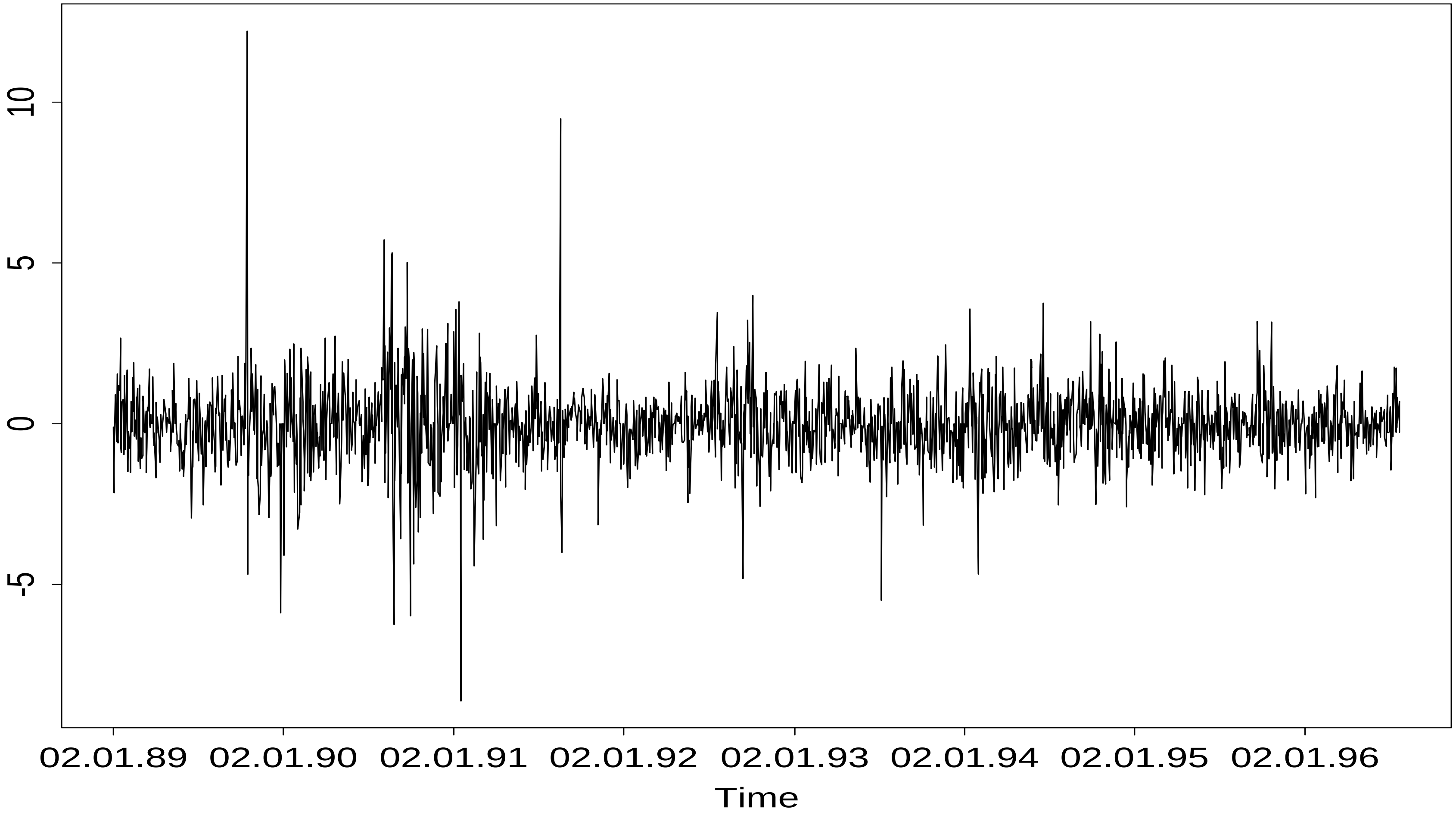
The Method

1. Construct the historical simulation data

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

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 $\mathbf{X}_{t+1}$.

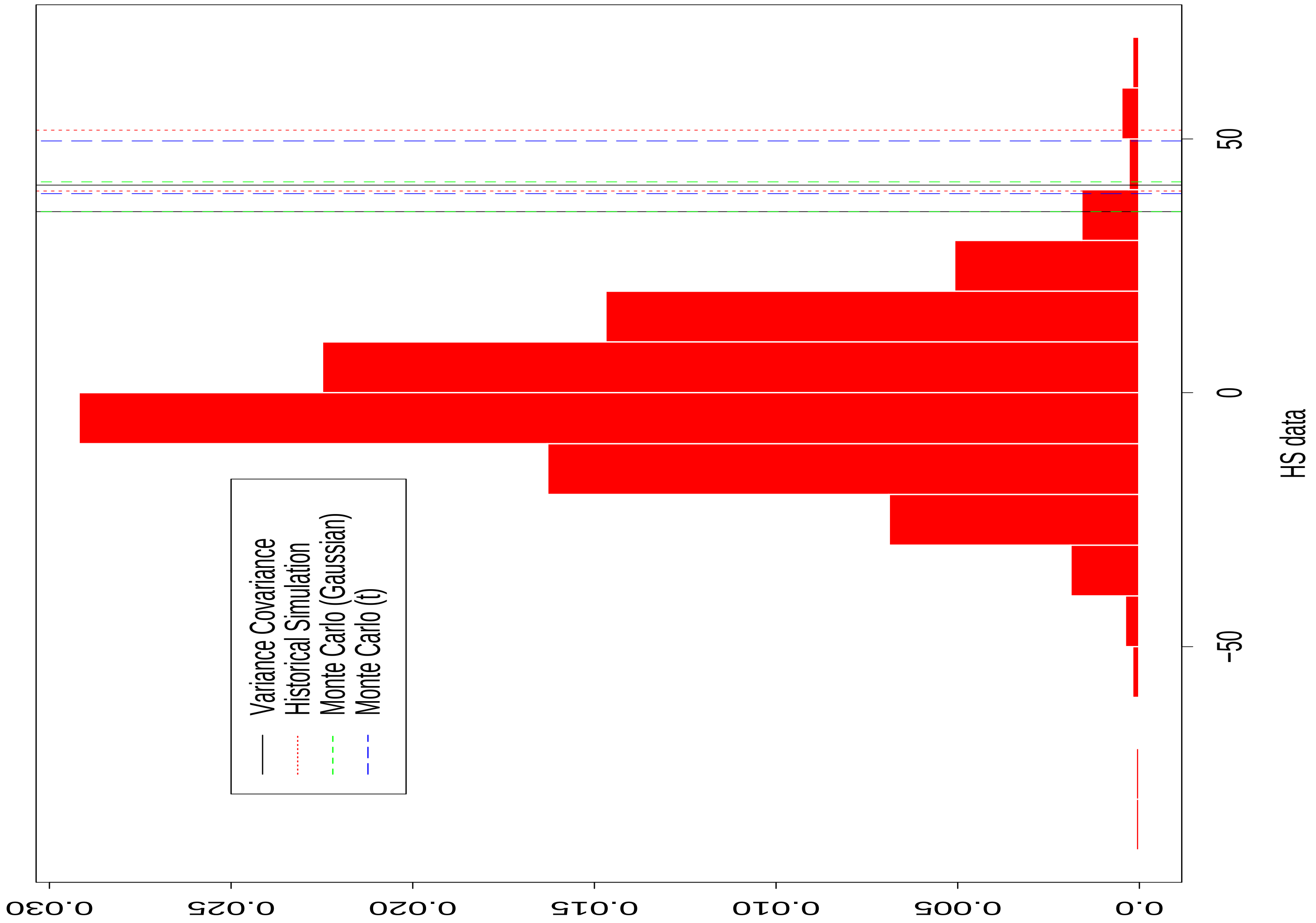
B4. An Example With BMW-SIEMENS Data

```
> 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)
```


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.

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

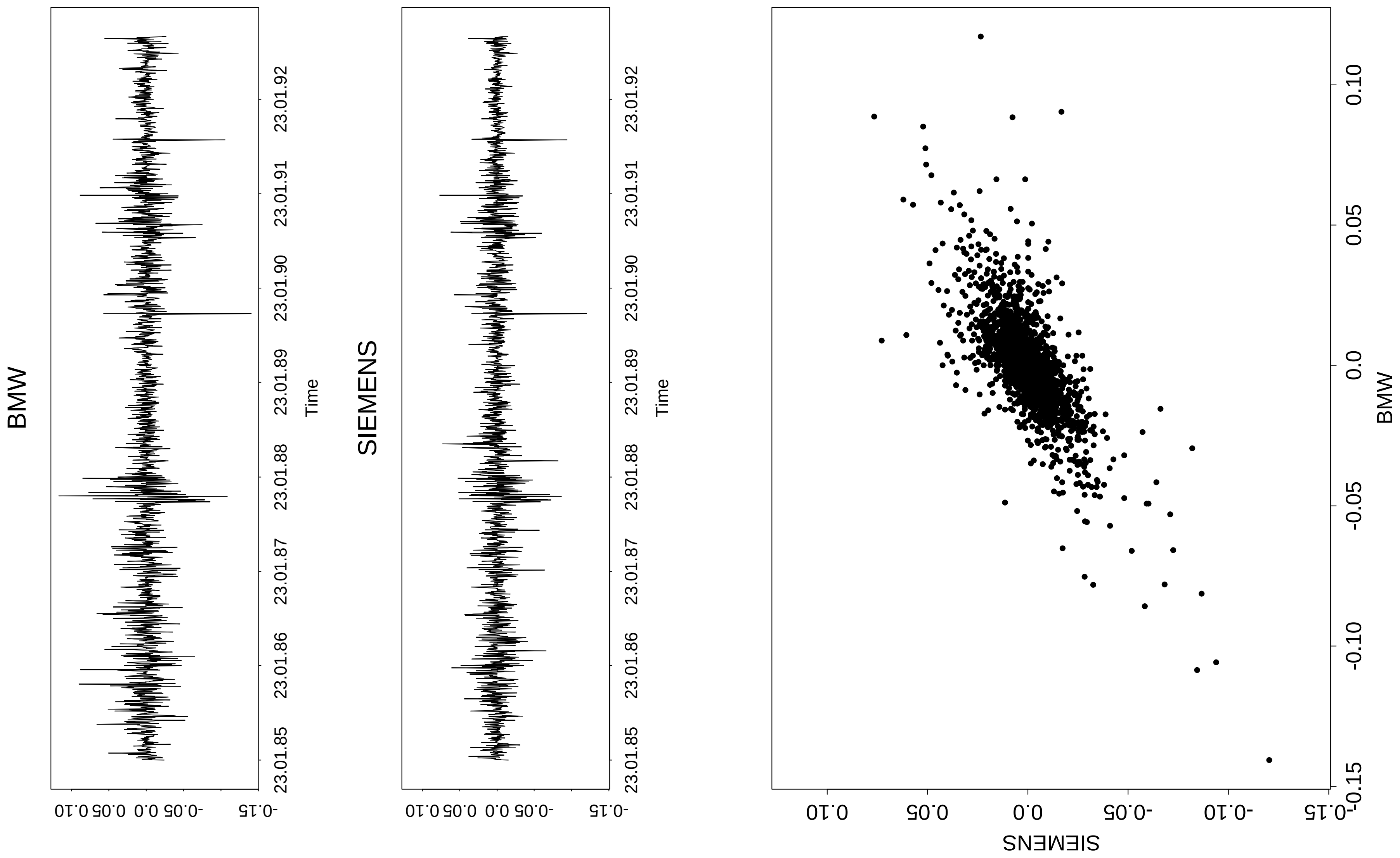
Assume we have data on risk factor changes $\mathbf{X}_1, \ldots, \mathbf{X}_n$. These might be daily (log) returns in context of market risk or longer interval returns in credit risk (e.g. monthly/yearly asset value returns). What are appropriate multivariate models?

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

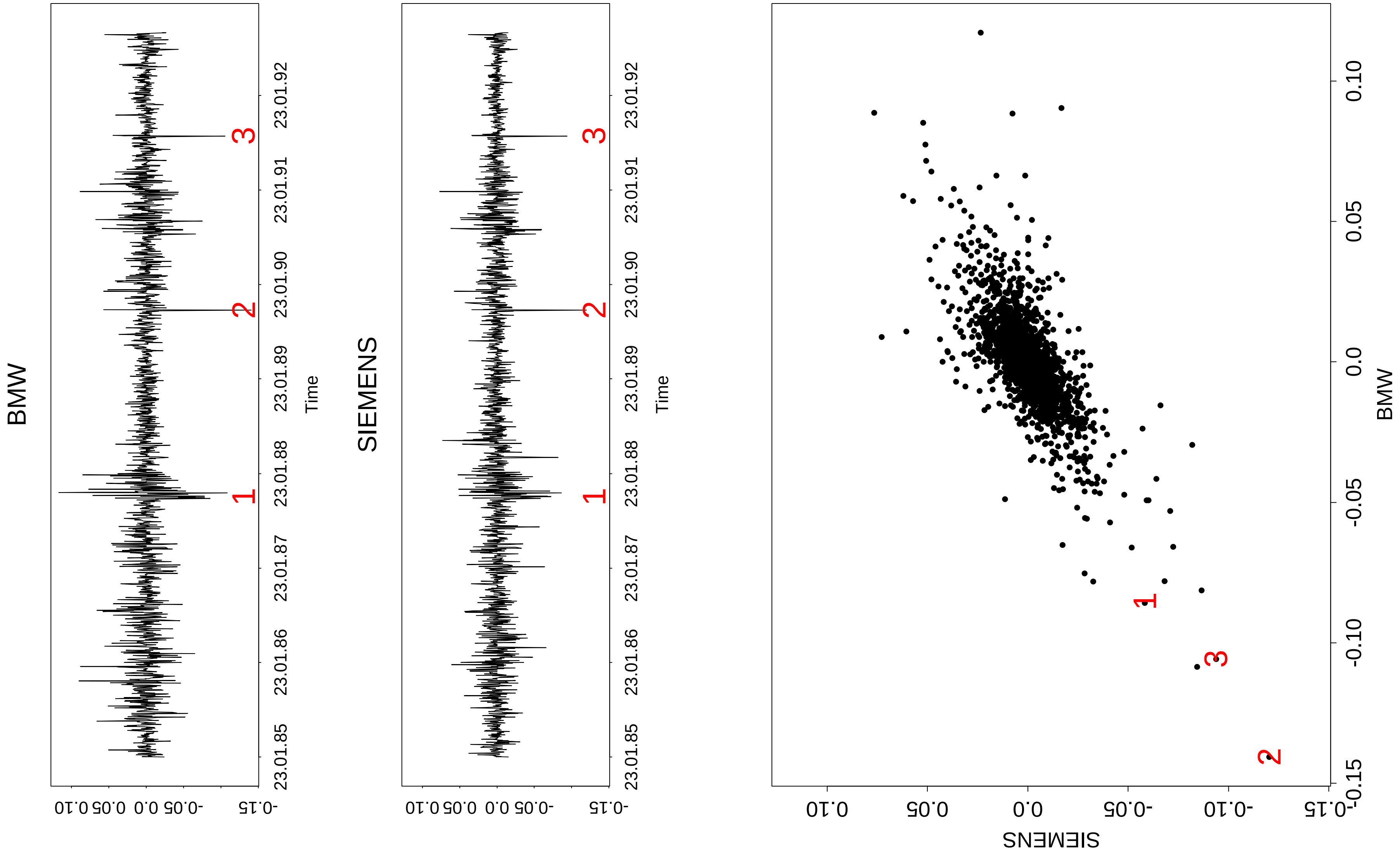
- Dynamic Models. In conditional approach we use multivariate time series models that allow us to make risk forecasts.

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

Bivariate Daily Return Data



Three Extreme Days



History


The Philadelphia Inquirer
Dow Dives 508.32 Points
In Panic on Wall Street
Navy blasts Iran oil site in retaliation
Billions lost in trading
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

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

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

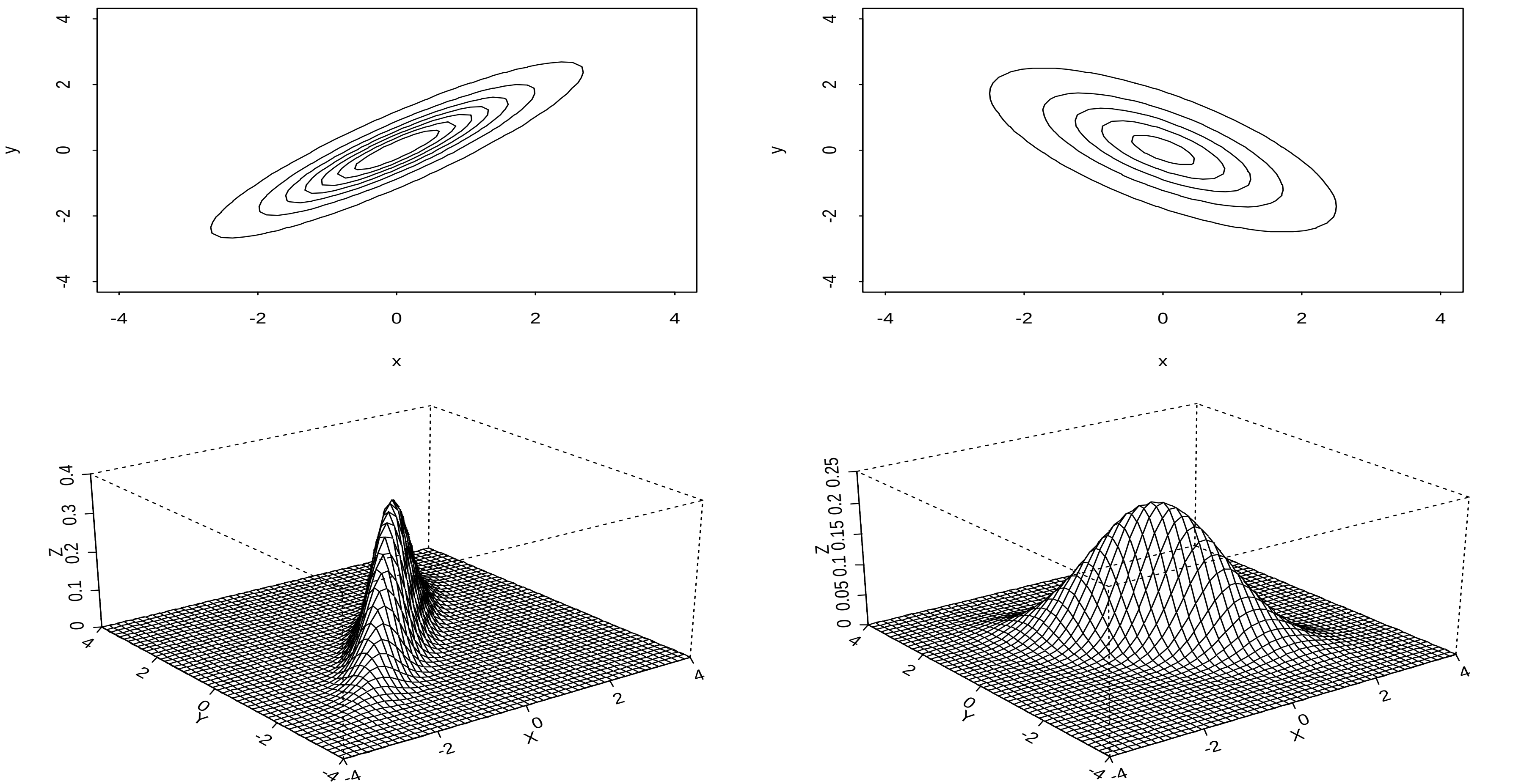
$$f(\mathbf{x}) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2}\right\},$$

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 $\mathrm{cov}(\mathbf{X}) = \Sigma$, 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 $(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) \sim \chi_d^2$ (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.

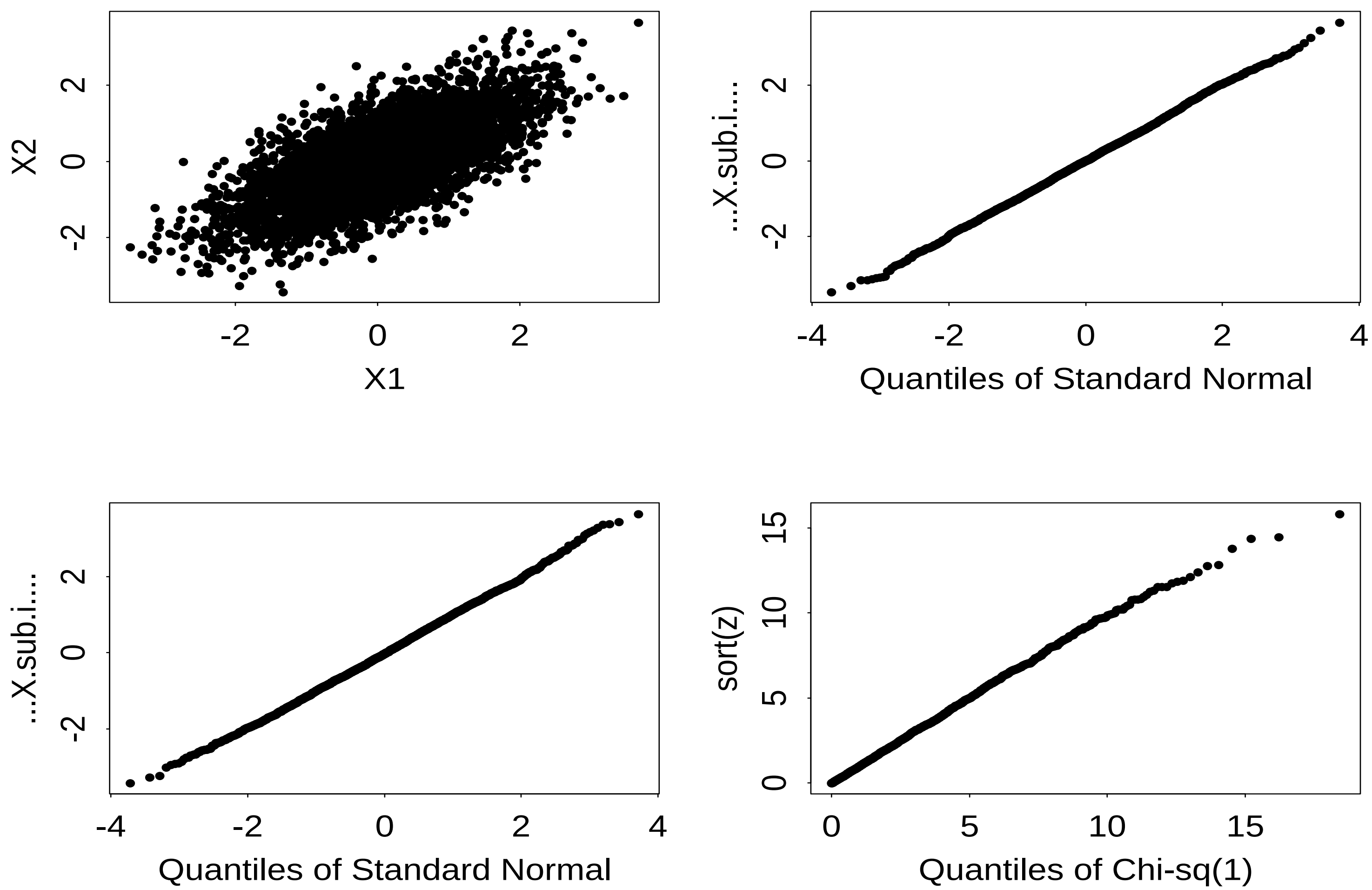
However, normality of the margins is not sufficient – we must test joint normality. To this end we calculate

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

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) $\operatorname{cov}(\mathbf{F}, \boldsymbol{\varepsilon}) = E((\mathbf{F} - E(\mathbf{F}))\boldsymbol{\varepsilon}') = 0.$

Remarks on Theory of Factor Models

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

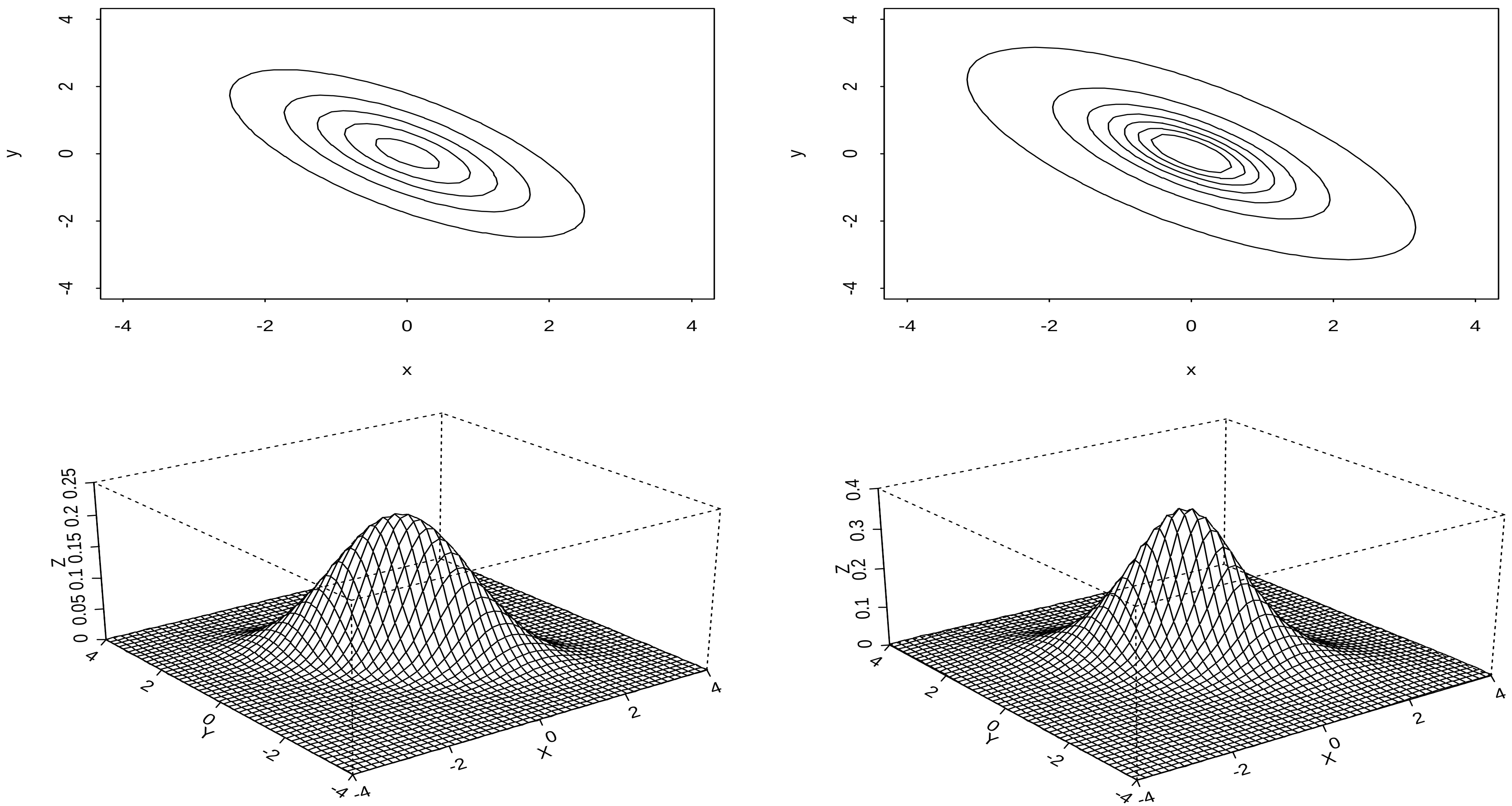
This has density

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

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.

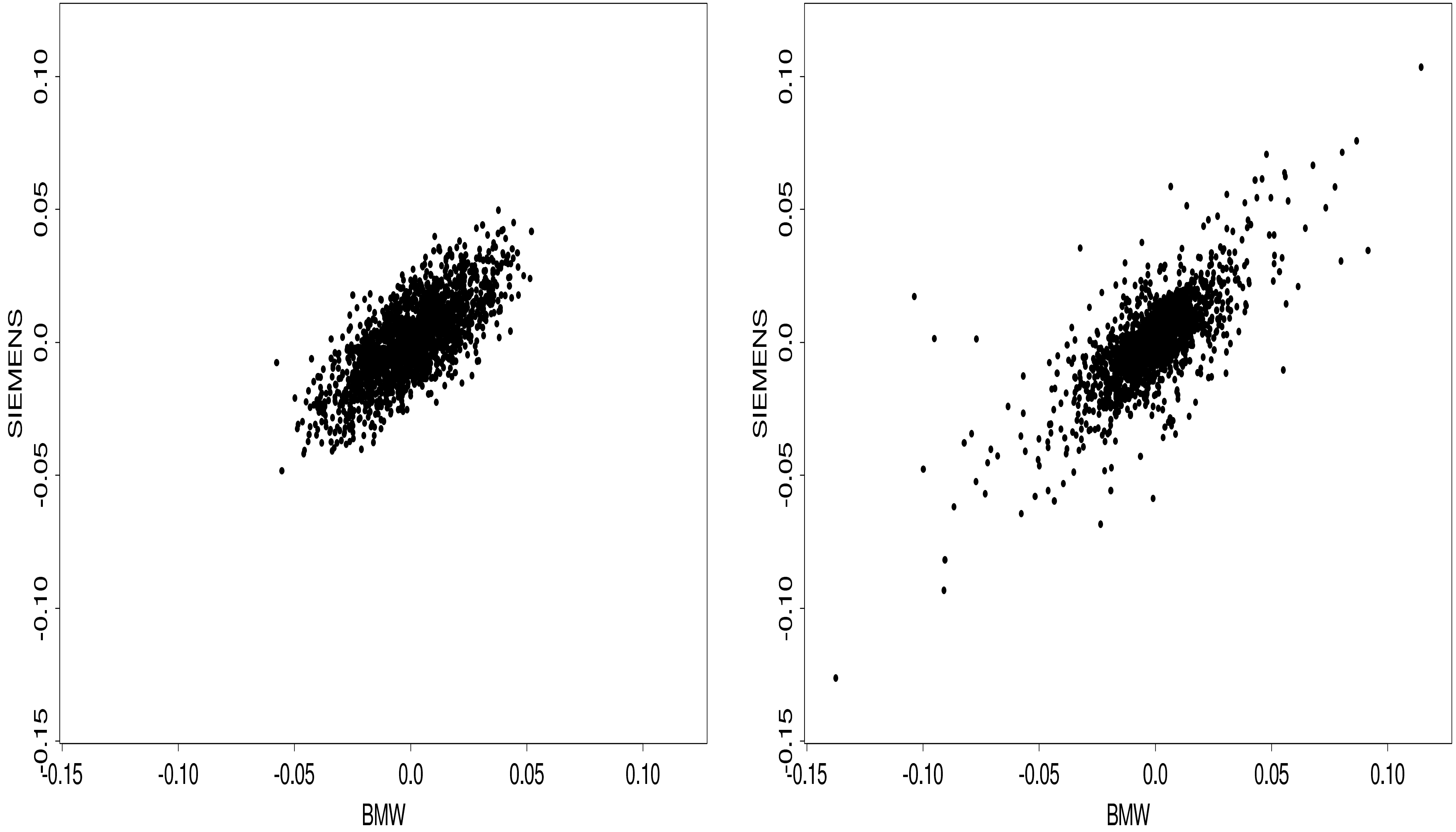
- If $\mathbf{X}$ has density f then $E(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{cov}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$, so that $\boldsymbol{\mu}$ and Σ are the mean vector and dispersion matrix respectively. For finite variances/correlations $\nu > 2$. Notation: $\mathbf{X} \sim t_d(\nu, \boldsymbol{\mu}, \Sigma)$.
- 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:

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

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., 2005]

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 $U'U = UU' = I_d$

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

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
  $\mathrm{VaR}_\alpha(Z_1+Z_2) \le \mathrm{VaR}_\alpha(Z_1)+\mathrm{VaR}_\alpha(Z_2)$, 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 $\varrho(\lambda Z) = \lambda \varrho(Z),\ \lambda \ge 0$,
  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., 2002] (elliptical distributions in RM)

E. Modelling Financial Time Series

1. Stylized Facts of Empirical Finance

2. Basics of Time Series Analysis

3. Classical Time Series Modelling with ARMA

4. Modelling Financial Time Series with ARCH and GARCH

5. Fitting GARCH Models to Financial Data

E1. Stylized facts of Financial Time Series

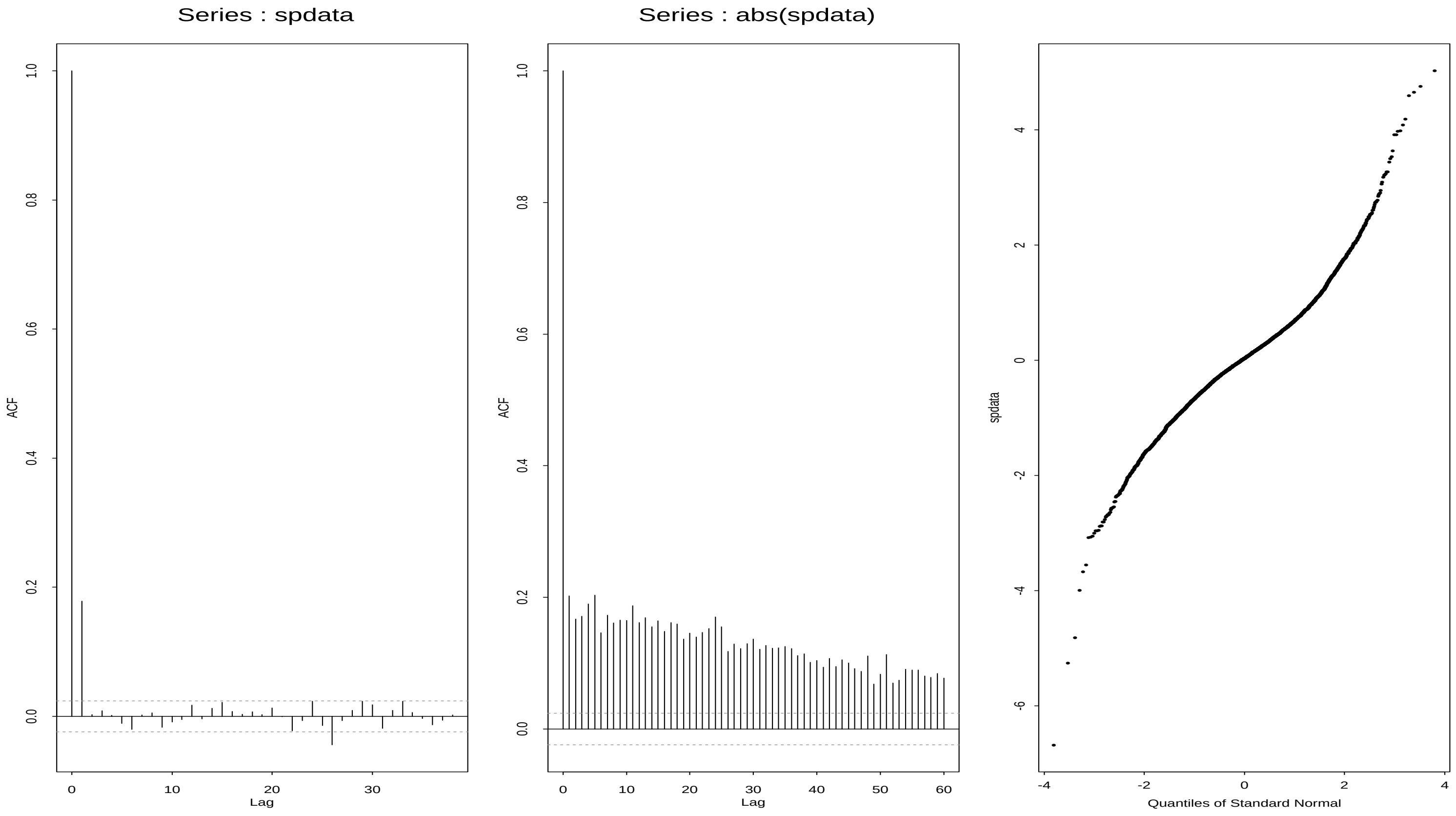
Consider discrete (e.g. daily) observations S_t of some asset price.
Let $X_t = (\log S_t - \log S_{t-1}) \approx (S_{t-1} - S_t)\,/S_{t-1}$, be the log returns.

A realistic model should reflect stylized facts of return series:

- Returns not iid but correlation low
- Absolute returns highly correlated
- Volatility appears to change randomly with time
- Returns are leptokurtic or heavy–tailed
- Extremes appear in clusters

Stylized Facts: Correlation, Heavy Tails

Correlograms of raw S&P data and absolute data,
and QQ–plot of raw data



Towards Models for Financial Time Series

We seek theoretical stochastic process models that can mimic these stylized facts. In particular, we require models that generate volatility clustering, since most of the other observations flow from this.

Econometricians have proposed a number of useful models including the ARCH/GARCH class.

We will concentrate on these (although there are alternatives such as discrete time stochastic volatility models).

To understand ARCH and GARCH it helps to briefly consider the classical family of ARMA models.

E2. Basics of Time Series Analysis

Stationarity

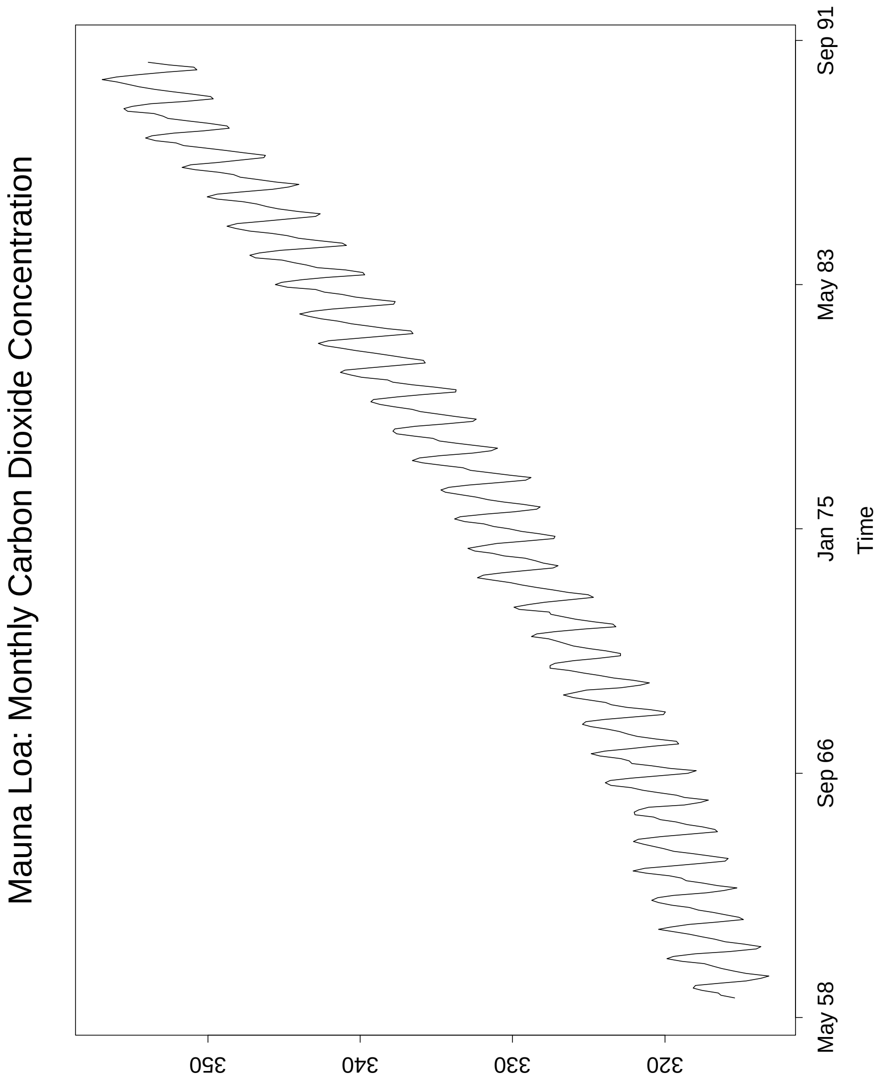
A time series $(X_t)_{t \in \mathbb{Z}}$ is strictly stationary if

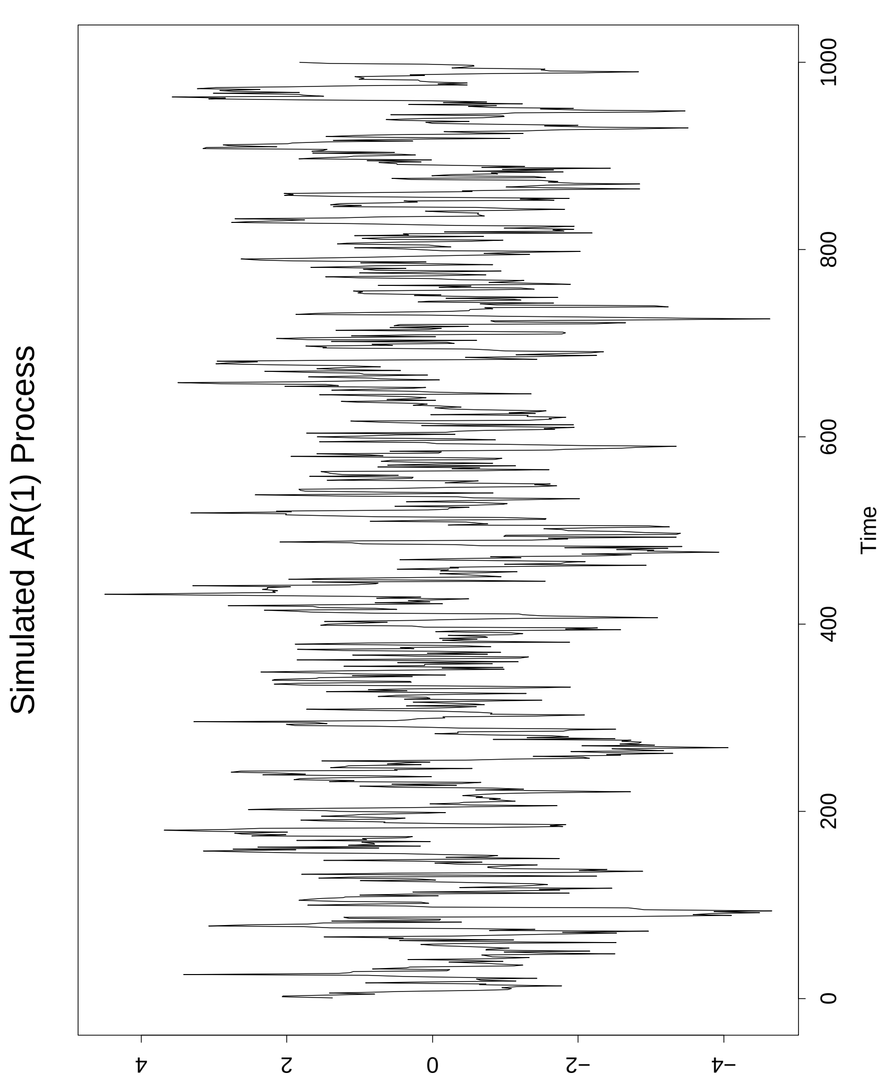
$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+h}, \ldots, X_{t_n+h})$$

for all $t_1, \ldots, t_n, h \in \mathbb{Z}$.

In particular this means that X_t has the same distribution for all $t \in \mathbb{Z}$, and this distribution is known as the stationary distribution (or marginal distribution).

Non-Stationary or Stationary?





Moments of a Stationary Time Series

For a strictly stationary time series $E(X_t)$ and $\mathrm{var}(X_t)$ must be constant for all t.
Moreover the autocovariance function defined by

$$\gamma(t,s) := \mathrm{cov}(X_t, X_s) = E\left((X_t - E(X_t))(X_s - E(X_s))\right).$$

must satisfy $\gamma(t,s) = \gamma(t+h, s+h)$ for all $t, s, h \in \mathbb{Z}$, which implies that covariance only depends on the separation in time of the observations $|t-s|$, also known as the lag.

A time series for which the first two moments are constant over time (and finite) and for which this condition holds, is known as covariance stationary, or second-order stationary.

The Autocorrelation Function

Rewrite the autocovariance function of a stationary time series as

$$\gamma(h) := \gamma(h,0) = \operatorname{cov}(X_h, X_0), \quad \forall h \in \mathbb{Z}.$$

Note, moreover that $\gamma(0) = \operatorname{var}(X_t),\ \forall t.$

The autocorrelation function is given by

$$\rho(h) := \rho(X_h, X_0), \quad \forall h \in \mathbb{Z}.$$

Clearly $\rho(h) = \gamma(h)/\gamma(0)$, in particular $\rho(0) = 1$.

We refer to $\rho(h), h = 1, 2, 3 \ldots$ as autocorrelations or serial correlations.

Time Domain and Frequency Domain

If we study dependence structure of a time series by analysing the autocorrelations we analyse (X_t) in the time domain.

There is an alternative approach based on Fourier analysis of the series, known as analysis in the frequency domain.

Most analysis of financial time series is done in the time domain, and we will restrict our attention to this.

An important instrument in the time domain is the correlogram, which gives empirical estimates of serial correlations.

The Correlogram

Given time series data $X_1, \ldots, X_n$ we calculate the sample autocovariances

$$\widehat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \overline{X})(X_{t+h} - \overline{X}) \quad \text{where} \quad \overline{X} = \sum_{t=1}^{n} X_t / n.$$

The sample autocorrelations are given by
$\widehat{\rho}(h) := \widehat{\gamma}(h)/\widehat{\gamma}(0),\ h = 0, 1, 2, \ldots.$

The correlogram is the plot $\{(h, \widehat{\rho}(h)),\ h = 0, 1, 2, \ldots\}$.

For many standard underlying processes, it can be shown that the $\widehat{\rho}(h)$ are consistent, and asymptotically normal estimators of the autocorrelations $\rho(h)$. (For very heavy-tailed processes, this theory does however break down.)

White Noise Processes

(These are processes with no appreciable dependence structure in the time domain.

A white noise process is a covariance stationary time series process whose autocorrelation function is given by

$$\rho(0) = 1, \quad \rho(h) = 0, \; h \neq 0.$$

That is, a process showing no serial correlation.

A strict white noise process is simply a process of iid random variables.

Not every white noise is a strict white noise. Plain ARCH and GARCH processes are in fact white noise processes!

E3. Classical ARMA Processes

Classical ARMA (autoregressive moving average) processes are constructed from white noise.

Let $(\varepsilon_t)_{t\in\mathbb{Z}}$ be a white noise process which has mean zero and finite variance σ_ε^2.

The (ε_t) form the innovations that drive the ARMA process.

Moving Average Process

These are defined as linear sums of the noise (ε_t).
(X_t) follows a MA(q) process if

$$X_t = \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t.$$

Autoregressive Process

These are defined by stochastic difference equations, or recurrence relations.

$(X_t)_{t\in\mathbb{Z}}$ follows a AR(p) process if for every t the rv X_t satisfies

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t,$$

where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise.

In order for these equations to define a covariance stationary causal process (depending only on past innovations) the coefficients ϕ_j must obey certain conditions.

ARMA Process

Autoregressive and moving average features can be combined to form ARMA processes.

$(X_t)_{t\in\mathbb{Z}}$ follows an ARMA(p,q) process if for every t the rv X_t satisfies

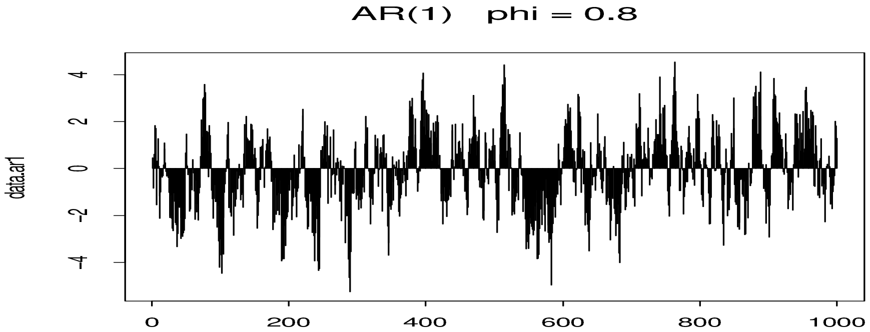
$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{17}$$

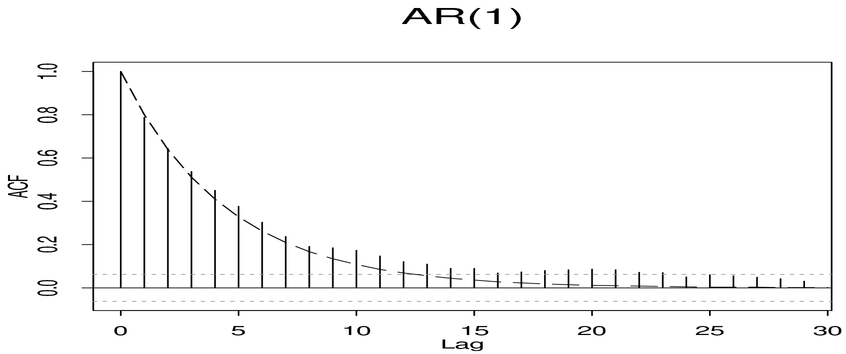
where $(\varepsilon_t)_{t\in\mathbb{Z}}$ is a white noise.

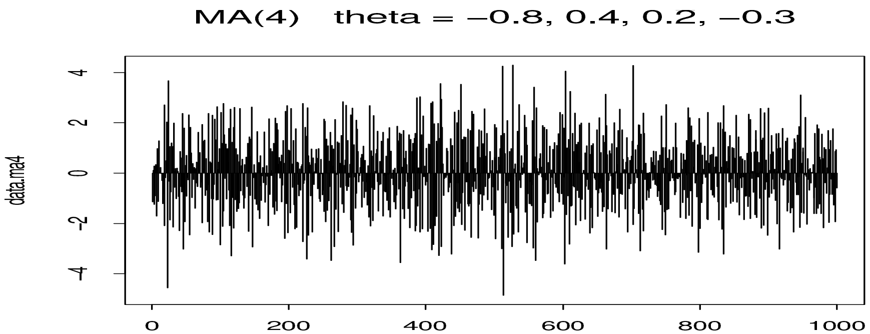
Again, there are conditions on the ϕ_j coefficients for these equations to define a covariance stationary causal process.

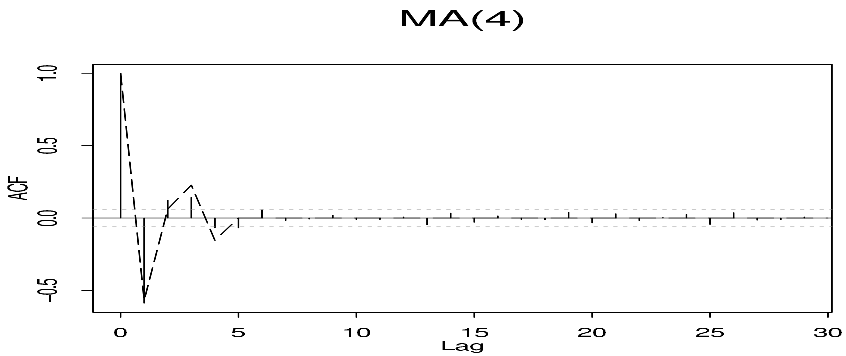
The autocorrelation functions of ARMA processes show a number of typical patterns, including exponential decay and damped sinusoidal decay.

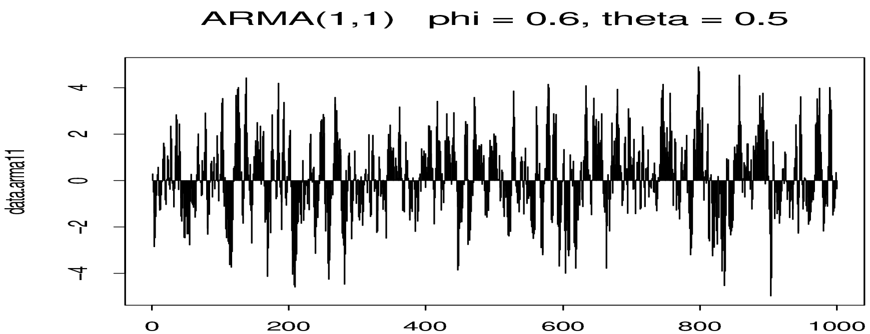
ARMA Examples

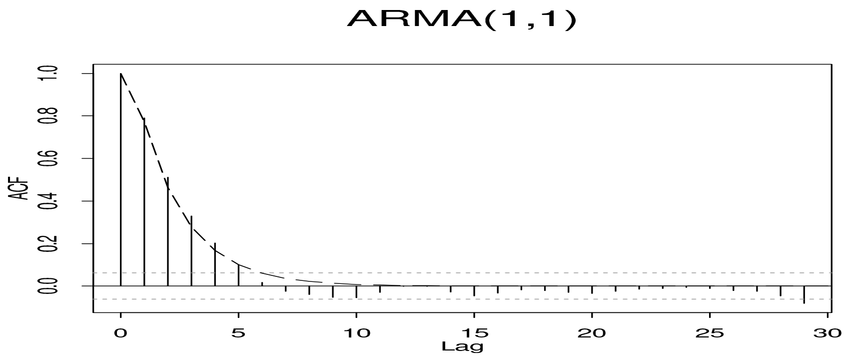












E4. Modelling Return Series with ARCH/GARCH

Let $(Z_t)_{t \in \mathbb{Z}}$ follow a strict white noise process with mean zero and variance one.

ARCH and GARCH processes $(X_t)_{t \in \mathbb{Z}}$ take general form

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z}, \tag{18}$$

where σ_t, the volatility, is a function of the history up to time $t-1$ represented by $\mathcal{F}_{t-1}$. Z_t is assumed independent of $\mathcal{F}_{t-1}$.

Mathematically, σ_t is $\mathcal{F}_{t-1}$-measurable, where $\mathcal{F}_{t-1}$ is the filtration generated by $(X_s)_{s \leq t-1}$, and therefore $\mathrm{var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$. Volatility is the conditional standard deviation of the process.

ARCH and GARCH Processes

(X_t) follows an ARCH(p) process if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \quad \text{with } \alpha_j > 0.$$

Intuition: volatility influenced by large observations in recent past.

(X_t) follows a GARCH(p,q) process (generalised ARCH) if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2, \quad \text{with } \alpha_j, \beta_k > 0. \qquad (19)$$

Intuition: more persistence is built into the volatility.

Stationarity and Autocorrelations

The condition for the GARCH equations to define a covariance stationary process with finite variance is that

$$\sum_{j=1}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k < 1.$$

ARCH and GARCH are technically white noise processes since

$$\begin{aligned} \gamma(h) = \operatorname{cov}(X_t, X_{t+h}) &= E(\sigma_{t+h} Z_{t+h} \sigma_t Z_t) \\ &= E(Z_{t+h}) E(\sigma_{t+h} \sigma_t Z_t) = 0. \end{aligned}$$

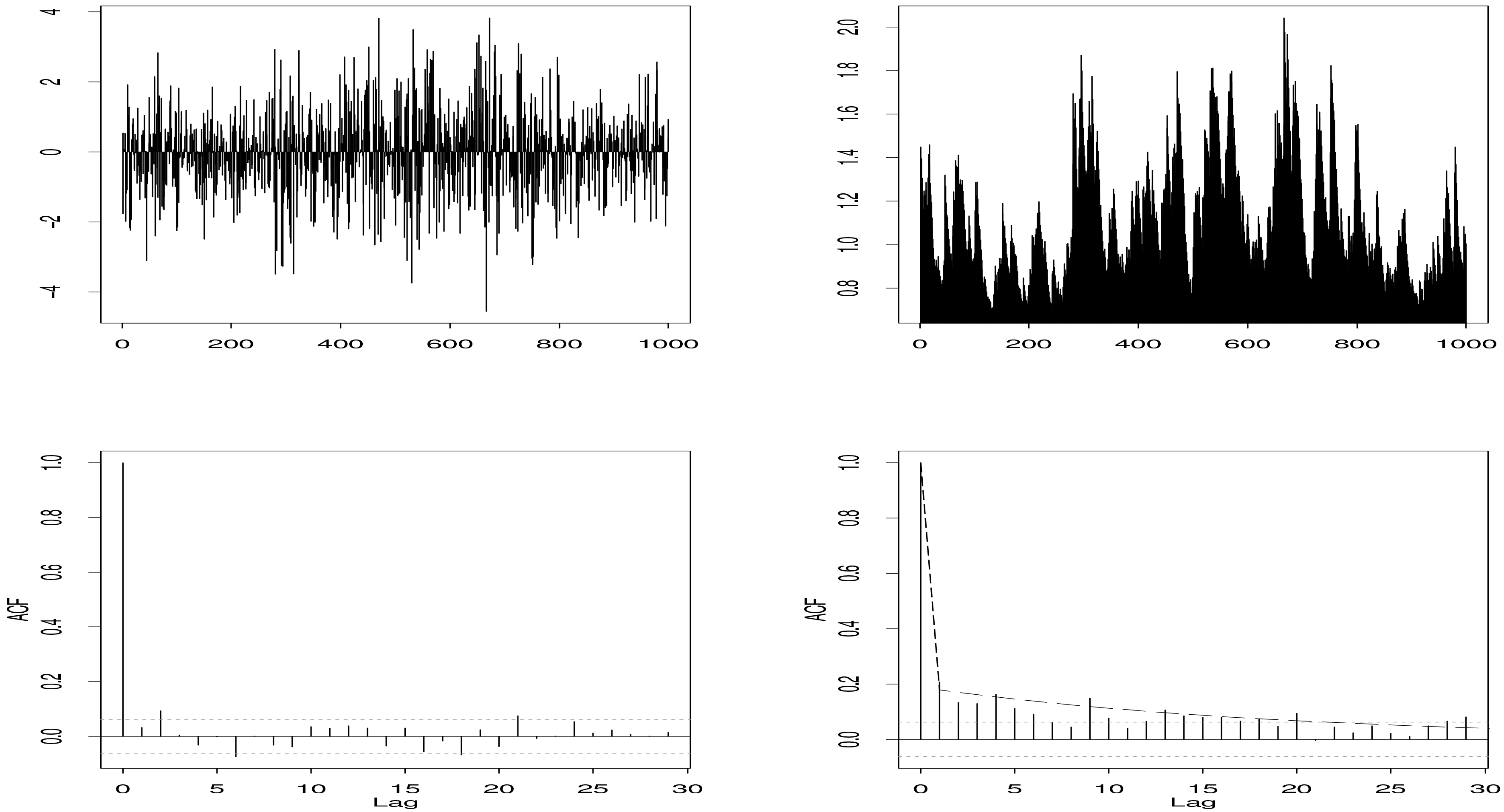
Absolute and Squared GARCH Processes

Although (X_t) is an uncorrelated process, it can be shown that the processes (X_t^2) and $(|X_i|)$ possess profound serial dependence.

In fact (X_t^2) can be shown to have a kind of ARMA-like structure.

A GARCH(1,1) model is like an ARMA(1,1) model for (X_t^2).

GARCH Simulated Example



Simulated realisation of series (upper left), volatility (upper right) together with correlograms of raw (lower left) and squared data.

Hybrid ARMA/GARCH Processes

Although changes in volatility are the most obvious feature of financial return series, there is sometimes some evidence of serial correlation at small lags. This can be modelled by

$$\begin{aligned} X_t &= \mu_t + \varepsilon_t \\ \varepsilon_t &= \sigma_t Z_t, \end{aligned} \tag{20}$$

where μ_t follows an ARMA specification, σ_t follows a GARCH specification, and (Z_t) is (0,1) strict white noise.

μ_t and σ_t are respectively the conditional mean and standard deviation of X_t given history to time $t-1$; they satisfy $E(X_t \mid \mathcal{F}_{t-1}) = \mu_t$ and $\mathrm{var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$.

A Simple Effective Model: AR(1)+GARCH(1,1)

For our purposes the following model will suffice.

$$\begin{aligned} \mu_t &= c + \phi(X_{t-1} - c)\,, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \left(X_{t-1} - \mu_{t-1}\right)^2 + \beta\sigma_{t-1}^2, \end{aligned} \tag{21}$$

with α_0, α_1, $\beta > 0$, $\alpha_1 + \beta < 1$ and $|\phi| < 1$ for a stationary model with finite variance.

This model is a reasonable fit for many daily financial return series, particularly under the assumption that the driving innovations are heavier-tailed than normal.

E5. Fitting GARCH Models to Financial Data

There are a number of possible methods, but the most common is maximum likelihood (ML), which is also a standard method of fitting ARMA processes to data.

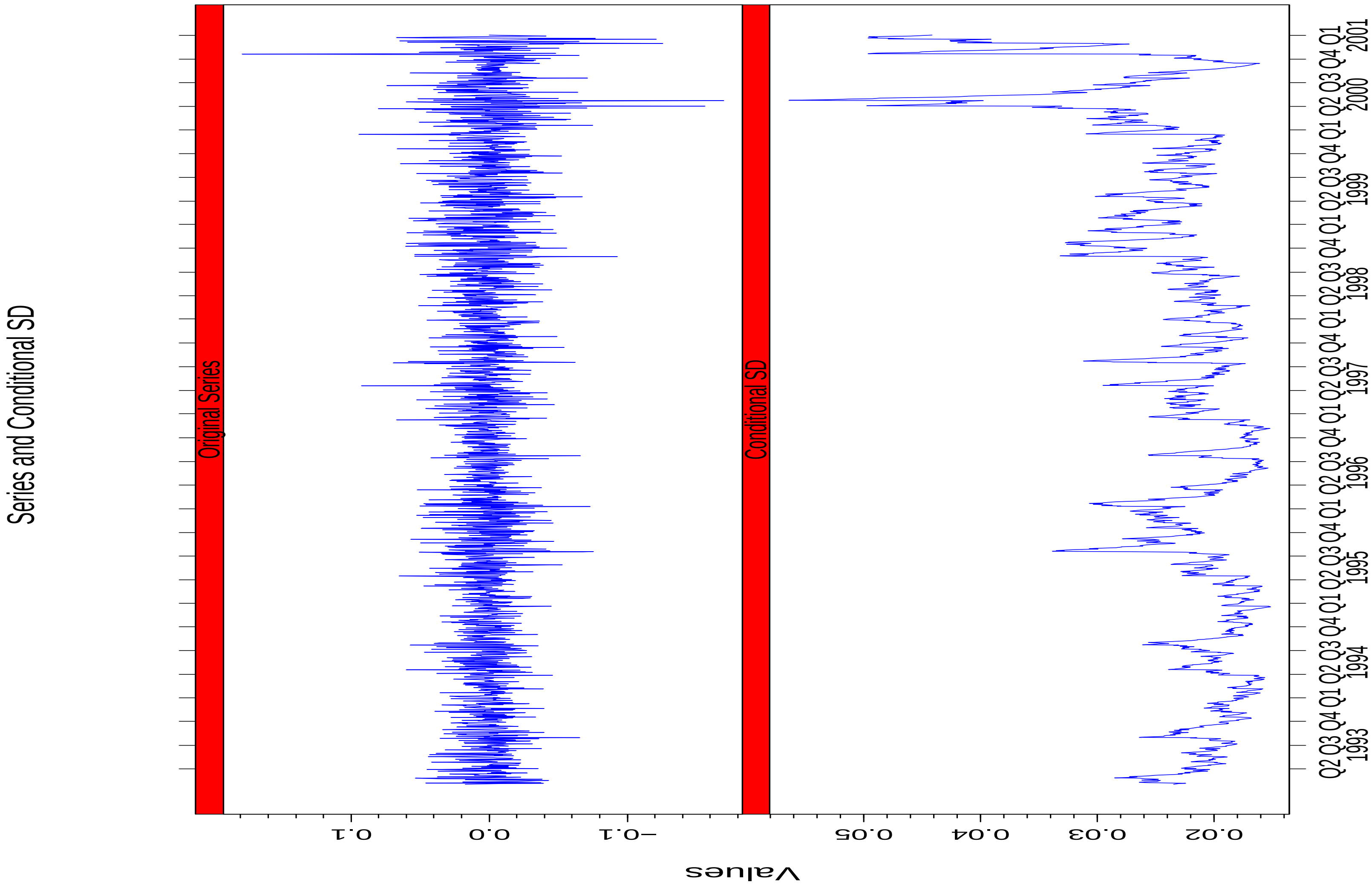
Possibilities:

- Assume (Z_t) are standard iid normal innovations and estimate GARCH parameters (α_j and β_k) by ML.
- Assume (Z_t) are (scaled) t_ν innovations and estimate GARCH parameters plus ν by ML.
- Make no distributional assumptions and estimate GARCH parameters by quasi maximum likelihood (QML). (Effectively uses Gaussian ML but calculates standard errors differently.)

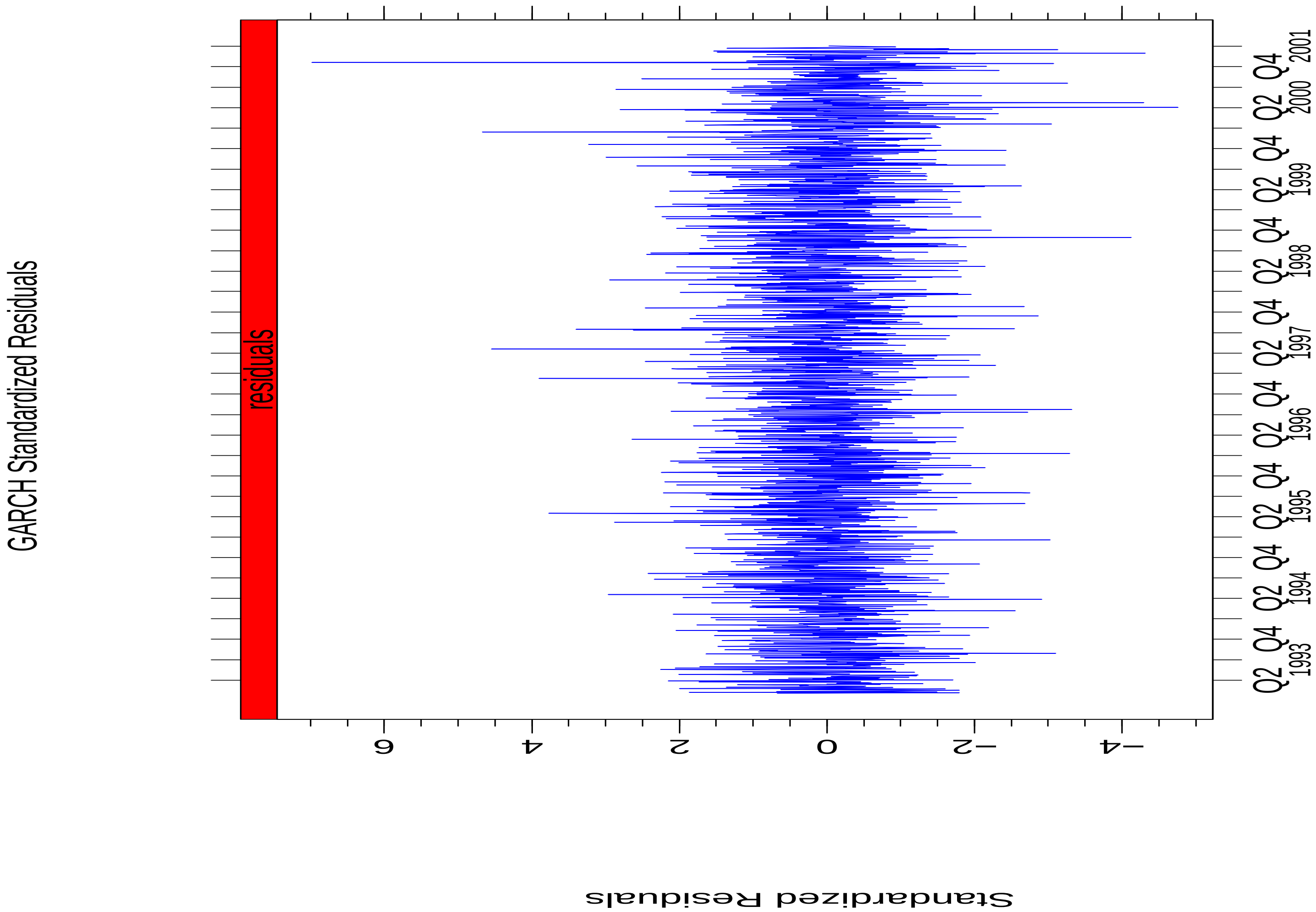
Example: Returns on Microsoft Share Price 93-01

```
Mean Equation: DJ30stock ~ ar(1)
Conditional Variance Equation:  ~ garch(1, 1)
Conditional Distribution:  t with estimated parameter 7.923343 and standard error 1.090603

Estimated Coefficients:
--------------------------------------------------------------
              Value Std.Error  t value  Pr(>|t|)
       C  1.247e-03 4.493e-04   2.7748 2.787e-03
   AR(1) -8.076e-03 2.261e-02  -0.3572 3.605e-01
       A  1.231e-05 4.580e-06   2.6877 3.628e-03
 ARCH(1)  5.876e-02 1.258e-02   4.6710 1.599e-06
GARCH(1)  9.188e-01 1.769e-02  51.9315 0.000e+00

Normality Test:
 Jarque-Bera P-value Shapiro-Wilk P-value
       451.5       0       0.9872  0.2192

Ljung-Box test for standardized residuals:
 Statistic P-value Chi^2-d.f.
     11.12  0.5184         12

Ljung-Box test for squared standardized residuals:
 Statistic P-value Chi^2-d.f.
     10.24  0.5948         12
```

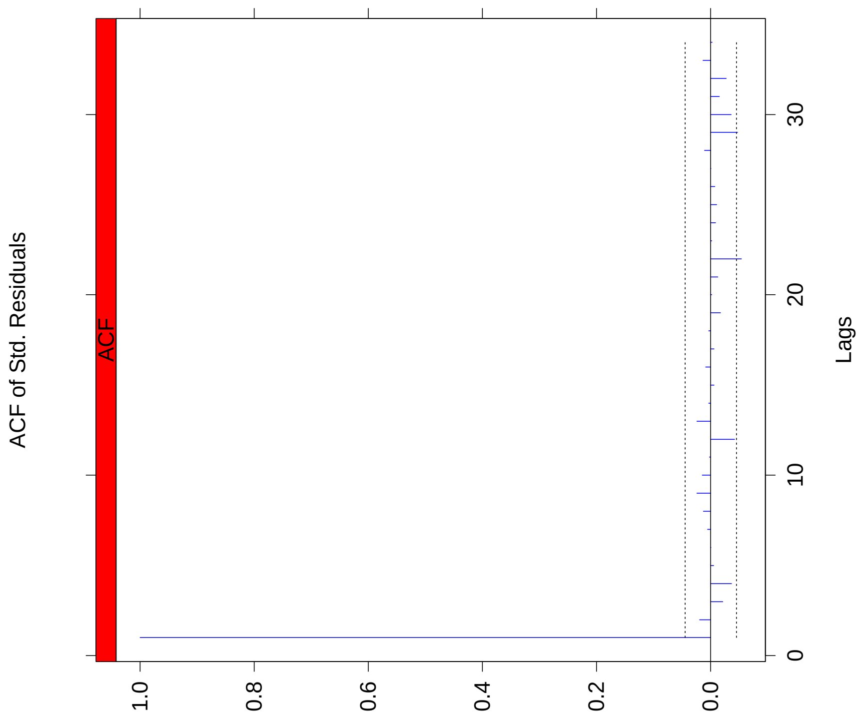
GARCH Analysis

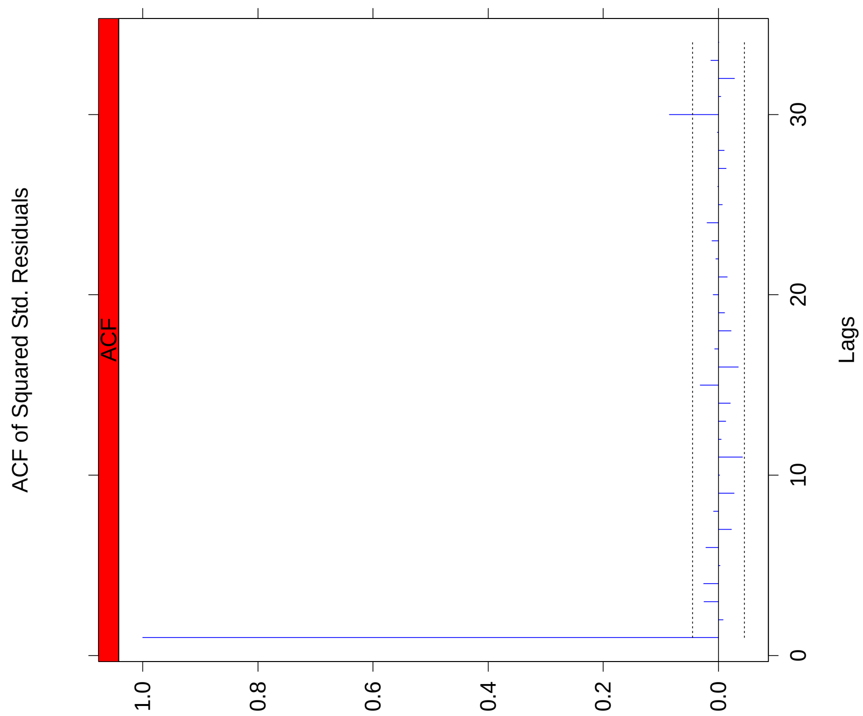


GARCH Analysis II

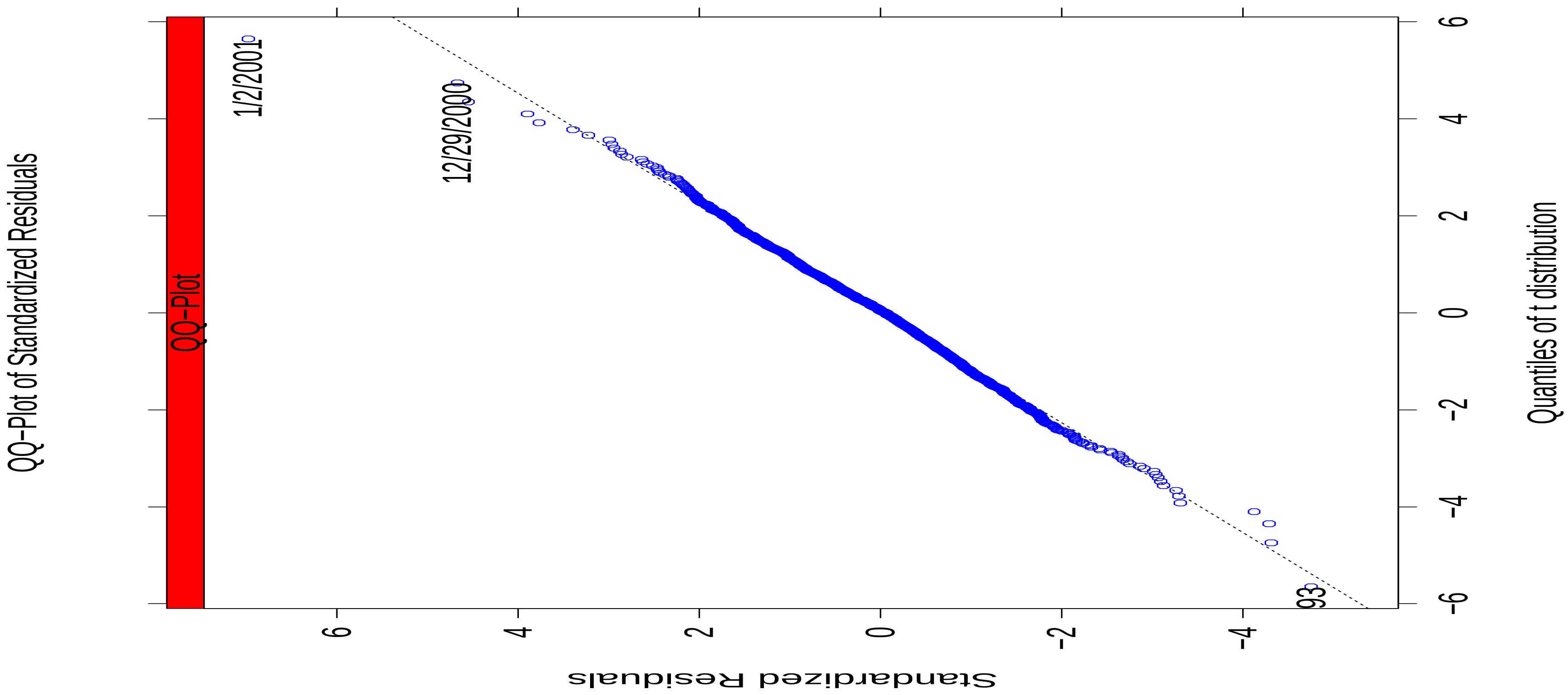


GARCH Analysis III





GARCH Analysis IV



F. The Dynamic Approach to Market Risk

1. The Conditional Problem

2. Backtesting

3. Longer Time Horizons

F1. Conditional Risk Measurement Revisited

In the dynamic or conditional problem interest centres on the distribution of the loss in the next time period $L_{t+1} = l_{[t]}(\mathbf{X}_{t+1})$, given $\mathcal{F}_t = \sigma\left\{(\mathbf{X}_s)_{s \leq t}\right\}$, the history of risk factor developments up to present.

Risk Measures based on the loss distribution, like VaR (6) and expected shortfall (7), are applied to the distribution of $L_{t+1} \mid \mathcal{F}_t$. We denote them as follows.

$$
\begin{aligned}
\mathrm{VaR}_\alpha^t &= F^{\leftarrow}_{[L_{t+1}|\mathcal{F}_t]}(\alpha), \\
\mathrm{ES}_\alpha^t &= E\left(L_{t+1} \mid L_{t+1} > \mathrm{VaR}_\alpha^t, \mathcal{F}_t\right).
\end{aligned}
$$

This problem forces us to consider the dynamics of the risk factor change time series and not just their long-term distribution.

The Single Risk Factor Problem

Suppose we consider an investment in a single stock or stock index, which has value V_t at time t. In terms of the log return for the next time period the loss follows (5) and is

$$L_{t+1} = -V_t \left(e^{X_{t+1}} - 1 \right) \approx -V_t X_{t+1} = L^{\Delta}_{t+1}.$$

We assume the time series (X_t) follows a model with structure

$$X_t = \mu_t + \sigma_t Z_t, \quad \mu_t, \sigma_t \in \mathcal{F}_{t-1}, \tag{22}$$

where (Z_t) are iid innovations with mean zero and unit variance. ARMA-GARCH models (20) and (21) provide example. This implies

$$L^{\Delta}_{t+1} = -V_t \mu_{t+1} - V_t \sigma_{t+1} Z_{t+1}. \tag{23}$$

Dynamic Risk Measures in 1-Risk-Factor Model

In this case the predictive distribution of X_{t+1} (or of $-L^{\Delta}_{t+1}$) is of the same type as the innovation distribution F_Z of the (Z_t).

$$F_{[X_{t+1}|\mathcal{F}_t]}(x) = P\left(X_{t+1} \leq x \mid \mathcal{F}_t\right) = F_Z\left(\frac{x - \mu_{t+1}}{\sigma_{t+1}}\right).$$

Risk measures applied to conditional distribution of L^{Δ}_{t+1} are:

$$\begin{aligned} \mathrm{VaR}^t_\alpha &= -V_t\mu_{t+1} + V_t\sigma_{t+1}q_\alpha(F_Z) \qquad (24) \\ \mathrm{ES}^t_\alpha &= -V_t\mu_{t+1} + V_t\sigma_{t+1}E(Z \mid Z > q_\alpha(F_Z)). \end{aligned}$$

They are linear transformations of the risk measures applied to the innovation distribution F_Z.

Risk Measure Estimation

Now assume we have data $X_{t-n+1}, \ldots, X_t$. We require a model that allows us to calculate:

1. Estimates/predictions $\widehat{\sigma}_{t+1}$ and $\widehat{\mu}_{t+1}$ of the volatility and conditional mean for the next time period;

2. Estimates $\widehat{q}_\alpha(F_Z)$ and $\widehat{\mathrm{ES}}_\alpha(Z)$ of risk measures applied to innovation distribution

In model with Gaussian innovations step (2) is immediate; $q_\alpha(F_Z) = \Phi^{-1}(\alpha)$ and $\mathrm{ES}_\alpha(Z) = \phi(\Phi^{-1}(\alpha))/(1-\alpha)$. In other models measures may depend on estimated parameters of innovation distribution.

Volatility and Conditional Mean Predictions

- GARCH Modelling. Fit ARMA-GARCH model (21) and use

$$\begin{aligned}\widehat{\mu}_{t+1} &= \widehat{c} + \widehat{\phi}(X_t - \widehat{c}) \\ \widehat{\sigma}^2_{t+1} &= \widehat{\alpha}_0 + \widehat{\alpha}_1(X_t - \widehat{\mu}_t)^2 + \widehat{\beta}_1\widehat{\sigma}^2_t,\end{aligned}$$

  where $\widehat{c}$, $\widehat{\phi}$, $\widehat{\alpha}_0$, $\widehat{\alpha}_1$, $\widehat{\beta}_1$ are parameter estimates and $\widehat{\mu}_t$ and $\widehat{\sigma}_t$ are also obtained from model.

- EWMA (RiskMetrics). We set $\mu_{t+1} = 0$ for simplicity and use a simple recursive scheme for volatility whereby $\widehat{\sigma}^2_{t+1} = (1-\lambda)X_t^2 + \lambda\widehat{\sigma}^2_t$, where λ is exponential smoothing parameter chosen by modeller (for example $\lambda = 0.94$).

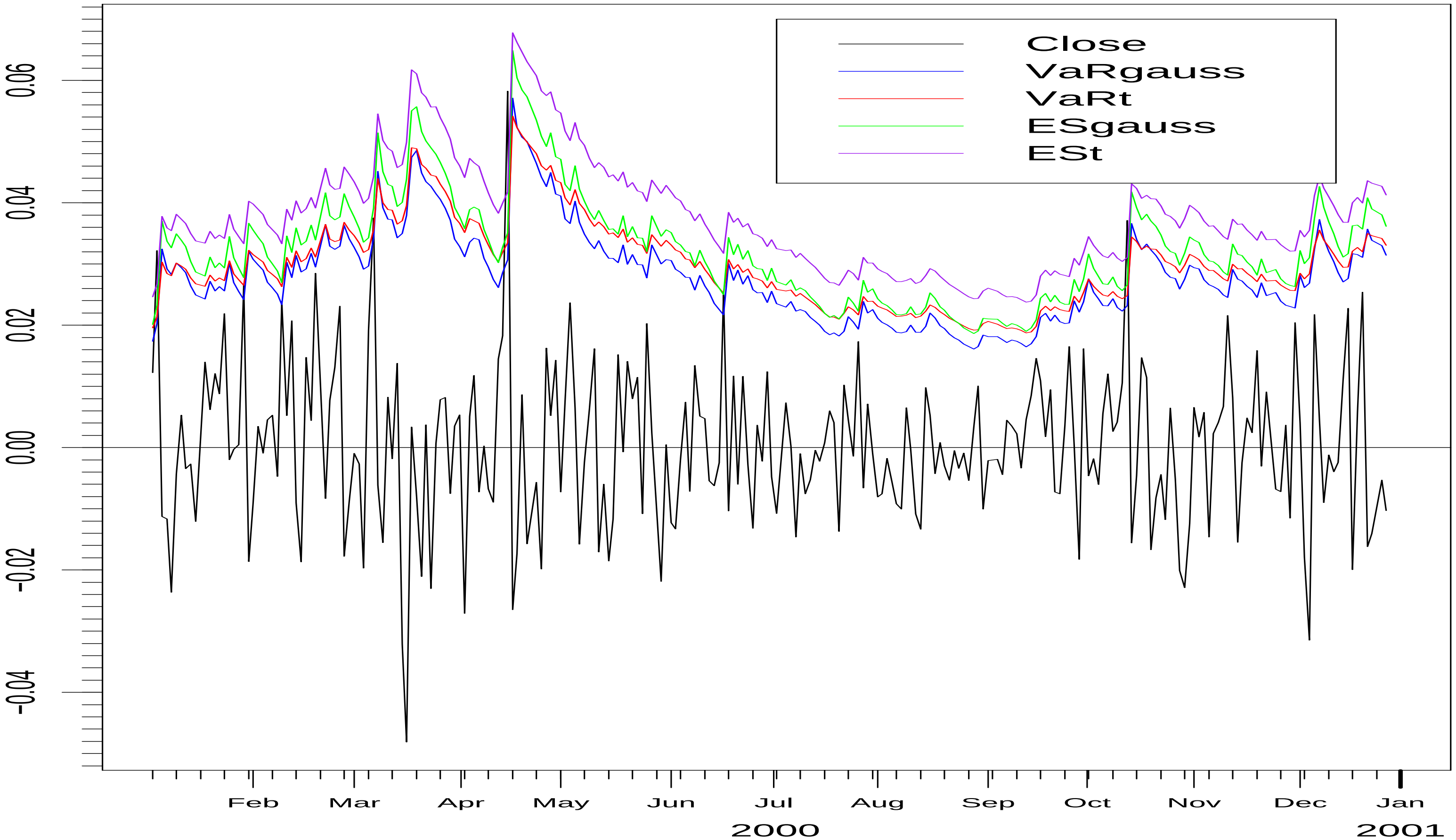
Risk Measures for the Scaled t Distribution

Write G_ν and g_ν for the df and density of standard univariate t. Suppose, for $\nu > 2$, that Z has a scaled t distribution so that $\sqrt{\nu/(\nu-2)}Z \sim t_\nu$. then

$$
\begin{aligned}
q_\alpha(F_Z) &= \sqrt{\frac{\nu-2}{\nu}} G_\nu^{-1}(\alpha) && (25)\\
\mathrm{ES}_\alpha(Z) &= \sqrt{\frac{\nu-2}{\nu}} \frac{g_\nu(G_\nu^{-1}(\alpha))}{1-\alpha} \left(\frac{\nu + (G_\nu^{-1}(\alpha))^2}{\nu-1} \right). && (26)
\end{aligned}
$$

When scaled t is used as innovation distribution in time series model ν is generally estimated. $\widehat{q}_\alpha(F_Z)$ and $\widehat{\mathrm{ES}}_\alpha(Z)$ are obtained by subtituting this estimate.

Dynamic Risk Measure Estimates (99%)



Daily losses for 1 unit invested in Dow Jones index in year 2000.

Ad Hoc Approach to Multiple Risk Factors

Recall the definition of the historical simulation data in (10). $\{\widetilde{L}_s = l_{[t]}(\mathbf{X}_s) \,:\, s = t-n+1, \ldots, t\}$. Suppose we fitted a univariate time series model to these data and used this to calculate conditional VaRs and expected shortfalls for L_{t+1}.

This can be thought of as attempting to estimate distribution of L_{t+1} given the information represented by $\mathcal{G}_t = \sigma\{(\widetilde{L}_s)_{s \le t}\}$. This is a subset of the information represented by $\mathcal{F}_t = \sigma\{(\mathbf{X}_s)_{s \le t}\}$.

In theory we have redefined the conditional problem and are calculating conditional estimates based on less information. In practice the estimates may often be quite good.

Genuinely multivariate methods discussed in Multivariate Time Series module.

F2. Backtesting

Let $I_t = 1_{\{L^{\Delta}_{t+1} > \mathrm{VaR}^t_\alpha\}}$ denote the indicator for a violation of the theoretical VaR of linearized loss on day $t+1$.

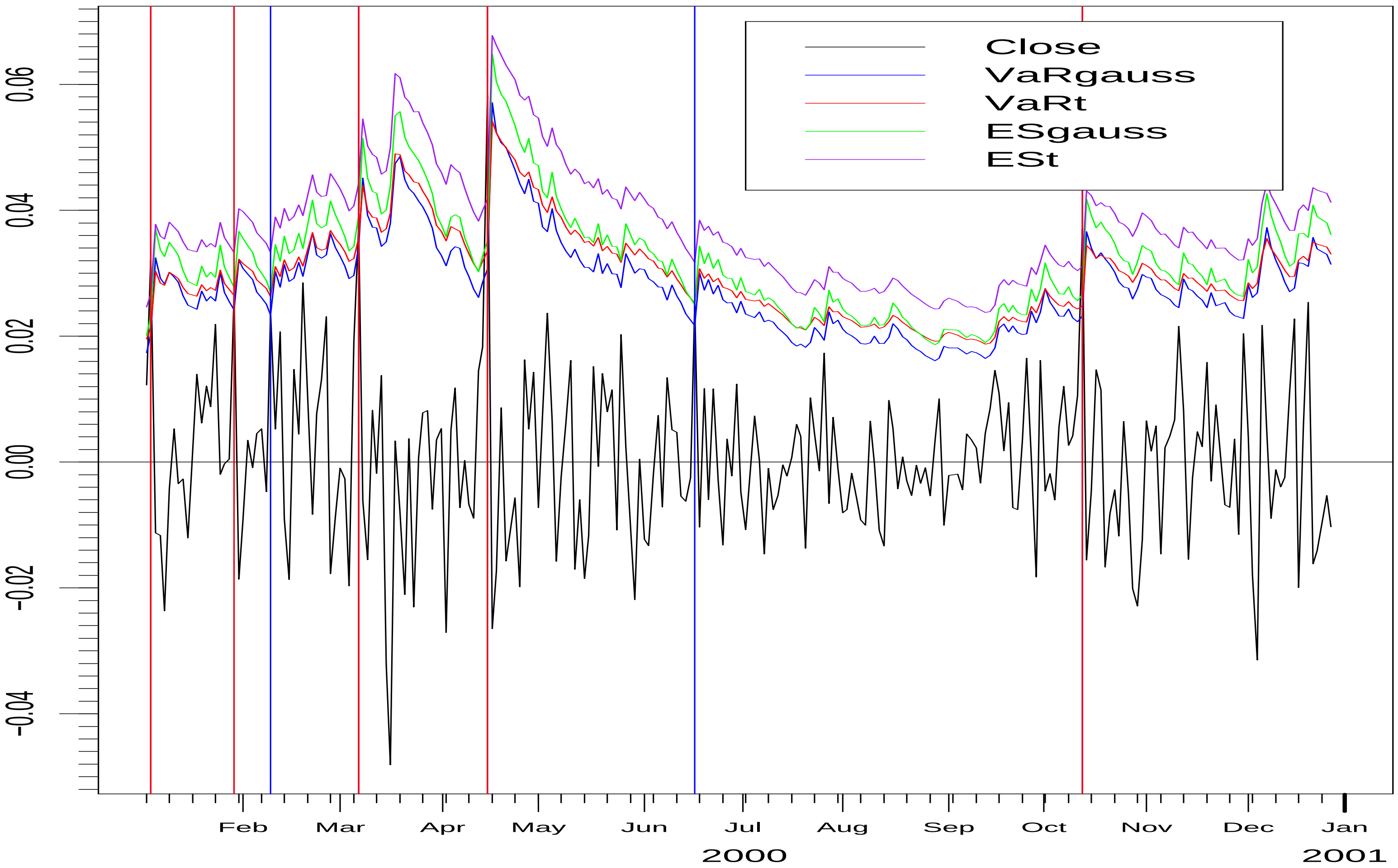
It follows from (23) and (24) that

$$I_t = 1_{\{Z_{t+1} > q_\alpha(F_Z)\}} \sim \text{Be}(1-\alpha);$$

moreover I_t and I_s are independent for $t \neq s$, since Z_{t+1} and Z_{s+1} are independent, so theoretical violation indicators are iid Bernoulli$(1-\alpha)$.

Of course we do not know VaR^t_α and will in practice look at the violation indicators $\widehat{I}_t = 1_{\left\{L^{\Delta}_{t+1} > \widehat{\mathrm{VaR}}^t_\alpha\right\}}$. We expect these to be roughly iid Bernoulli$(1 - \alpha)$.

Violations of 99% VaR Estimates



Violation Count Tables and Binomial Tests

Quantile	Method	S&P $n = 7414$	DAX $n = 5146$
95%	Expected	371	257
95%	GARCH (Normal)	384 (0.25)	238 (0.11)
95%	GARCH (t)	404 (0.04)	253 (0.41)
99%	Expected	74	51
99%	GARCH (Normal)	104 (0.00)	74 (0.00)
99%	GARCH (t)	78 (0.34)	61 (0.11)

Expected and observed violation counts for VaR estimates for two market indices obtained from GARCH modelling (Gaussian and scaled t innovations). Methods use last 1000 data values for each forecast. p-value for binomial test given in brackets.

Expected Shortfall to Quantile Ratios

In general we have $\text{ES}_\alpha^t / \text{VaR}_\alpha^t \approx E\left(Z \mid Z > q_\alpha(F_Z)\right)/q_\alpha(F_Z)$ and it is interesting to look at the typical magnitude of such ratios.

α	0.95	0.99	0.995	$q \to 1$
t_4	1.5	1.39	1.37	1.33
Normal	1.25	1.15	1.12	1.00

Conclusion.

If GARCH with Gaussian innovations has a tendency to underestimate VaR at higher levels ($\alpha \geq 99\%$), then it will have an even more pronounced tendency to underestimate expected shortfall.

Heavy-tailed innovation distributions required to estimate ES_α^t.

F3. Risk Measure Estimates for h-Period Loss

Suppose we require conditional VaR and ES estimates for the loss distribution over a 1-week, 2-week, 1-month or other period. There are two possible strategies:

- Base analysis on loss data for appropriate period length, e.g. non-overlapping weekly or fortnightly losses. Advantage: estimates obtained immediately from model. Disadvantage: reduction in quantity of data.

- Infer risk measure estimates for longer period from a model fitted to higher frequency returns, typically daily. Advantage: more data. Disadvantage: scaling behaviour of reasonable models not well understood.

The Single Risk Factor Problem

Let (S_t) and (X_t) denote daily asset value and log-return process respectively. The h period log-return at time t is given by

$$X_{t+h}^{(h)} = \log\left(\frac{S_{t+h}}{S_t}\right) = X_{t+1} + \cdots + X_{t+h}.$$

If our position in asset has value V_t at time t our loss is $L_{t+h}^{(h)} = -V_t(\exp(X_{t+h}^{(h)} - 1)$ and we wish to estimate risk measures

$$\begin{aligned}
\mathsf{VaR}_\alpha^{t,h} &= F^{\leftarrow}_{\left[L_{t+h}^{(h)} \mid \mathcal{F}_t\right]}(\alpha), \\
\mathsf{ES}_\alpha^{t,h} &= E\left(L_{t+h}^{(h)} \mid L_{t+h}^{(h)} > \mathsf{VaR}_\alpha^{t,h}, \mathcal{F}_t\right).
\end{aligned}$$

A Monte Carlo Approach

1. Fit model to daily returns $X_{t-n+1}, \ldots, X_t$.

2. By simulating independent innovations construct many future paths $X_{t+1}, \ldots, X_{t+h}$.

3. Calculate $X_{t+h}^{(h)}$ and $L_{t+h}^{(h)}$ for each path.

4. Estimate risk measures by empirical quantile and shortfall estimation techniques applied to simulated values of $L_{t+h}^{(h)}$.

Alternative Method. Estimate one-step risk measures $\widehat{\mathrm{VaR}}_{\alpha}^{t}$ and $\widehat{\mathrm{ES}}_{\alpha}^{t}$. Scale these estimates by square–root–of–time $\sqrt{h}$ to obtain estimates for h-day period (valid for iid normally distributed returns).

Some Empirical Results

	S&P	DAX	BMW
$h = 10$; length of test	7405	5136	5136
0.95 Quantile			
Expected	370	257	257
Simulation Method (h-day)	403	249	231
Square–root–of–time	623	318	315
0.99 Quantile			
Expected	74	51	51
Simulation Method (h-day)	85	48	53
Square-root-of-time	206	83	70

Conclusion. Square root of time scaling does not seem sophisticated enough! Note that formal statistical testing difficult because of overlapping returns.

G. Multivariate Financial Time Series

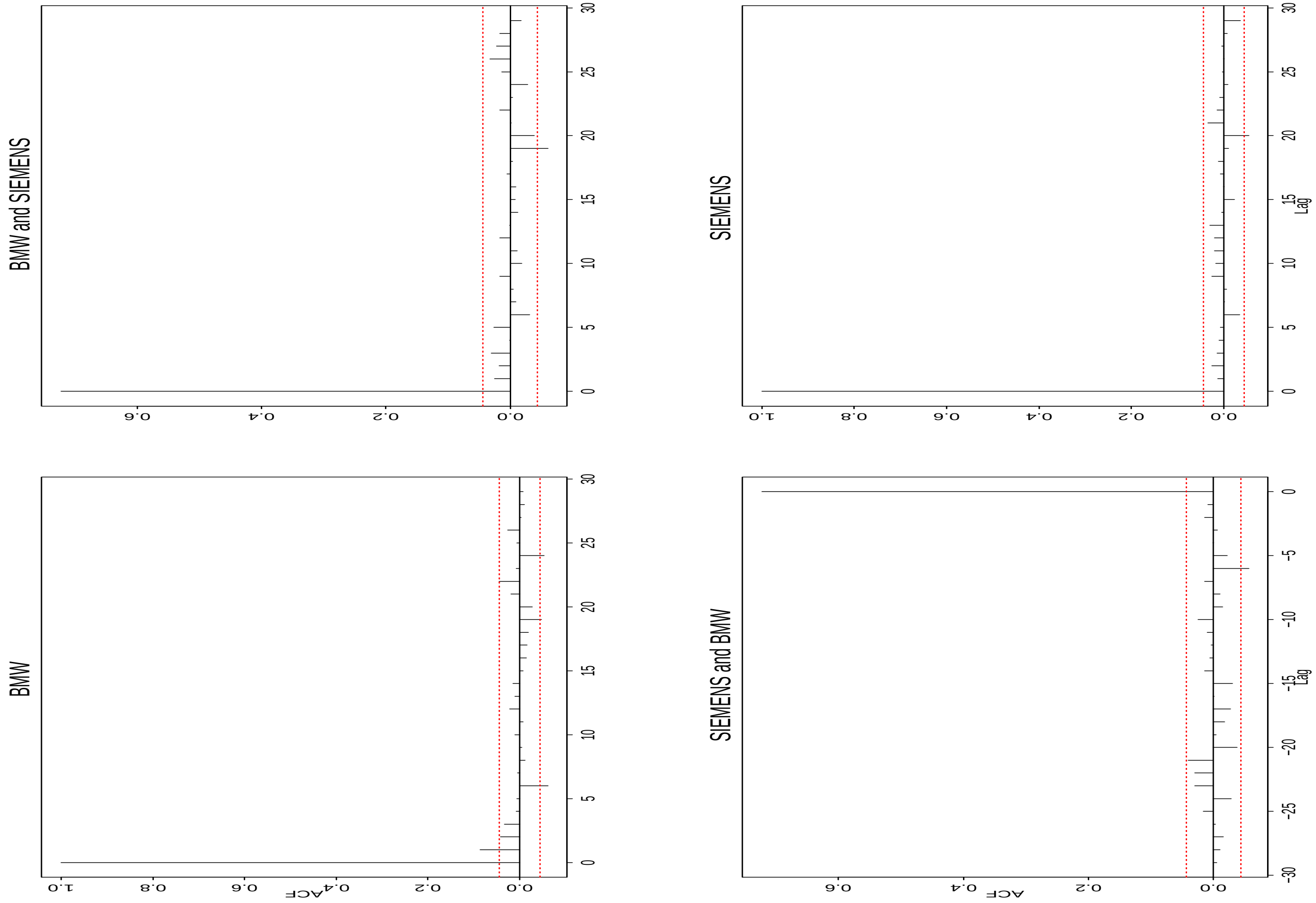
1. Stylized Facts of Multivariate Risk Factors
2. Basics of Multivariate Time Series
3. General Multivariate GARCH Model
4. Constant Conditional Correlation GARCH
5. The DVEC GARCH Model
6. Risk Management with Multivariate GARCH

G1. Multivariate Stylized Facts

We have observed a number of univariate stylized facts. These may be augmented by the following multivariate observations.

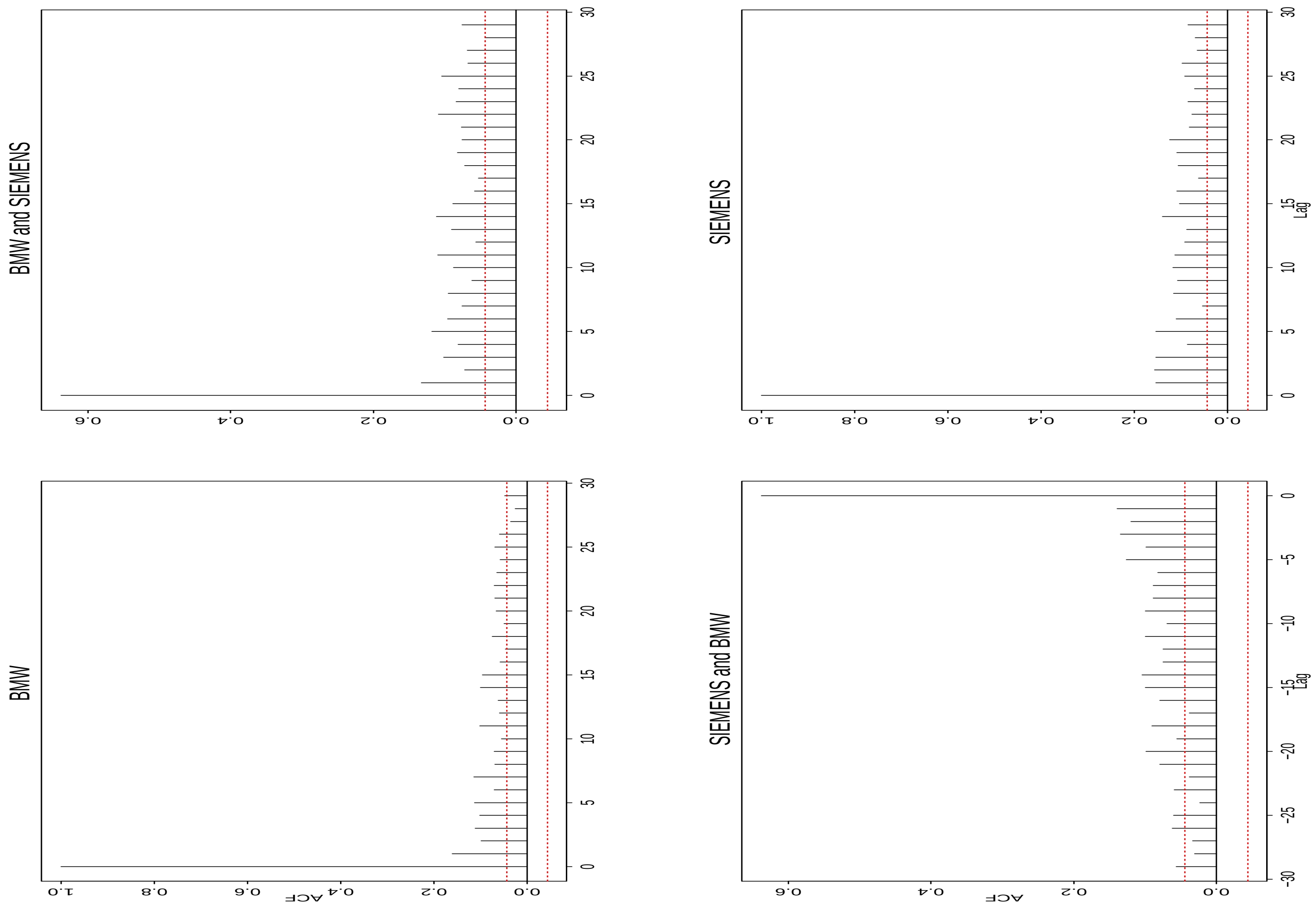
- Return series show little cross correlation, except at lag 0.
- Cross-correlogram of absolute or squared returns show profound cross-correlations.
- "Correlations appear to be higher in stress periods than normal periods"
- Extreme moves of many financial assets are synchronous.

Cross-Correlogram of BMW-Siemens Data



2000 values from period 1985-1993

Cross-Correlogram of Absolute BMW-Siemens Data



2000 values from period 1985-1993

G2. Basics of Multivariate Time Series

Stationarity. (Recall Univariate Case)

A multivariate time series $(\mathbf{X}_t)_{t \in \mathbb{Z}}$ is strictly stationary if

$$(\mathbf{X}'_{t_1}, \ldots, \mathbf{X}'_{t_n}) \stackrel{\mathrm{d}}{=} (\mathbf{X}'_{t_1+h}, \ldots, \mathbf{X}'_{t_n+h})$$

for all $t_1, \ldots, t_n, h \in \mathbb{Z}$.

In particular this means that $\mathbf{X}_t$ has the same multivariate distribution for all $t \in \mathbb{Z}$, and this distribution is known as the stationary distribution (or marginal distribution).

Moments of a Stationary Multivariate Time Series

For a strictly stationary multivariate time series $E(\mathbf{X}_t)$ and $\mathrm{cov}(\mathbf{X}_t)$ must be constant for all t.
Moreover the covariance matrix function defined by

$$\Gamma(t,s) := \mathrm{cov}(\mathbf{X}_t, \mathbf{X}_s) = E\left((\mathbf{X}_t - E(\mathbf{X}_t))(\mathbf{X}_s - E(\mathbf{X}_s))'\right).$$

must satisfy $\Gamma(t,s) = \Gamma(t+h, s+h)$ for all $t, s, h \in \mathbb{Z}$

A multivariate time series for which the first two moments are constant over time (and finite) and for which this condition holds, is known as covariance stationary, or second-order stationary.

(Recall univariate case)

Covariance Matrix Function of Stationary T.S.

We may rewrite the covariance matrix function of a stationary time series as

$$\Gamma(h) := \Gamma(h, 0) = \mathrm{cov}(\mathbf{X}_h, \mathbf{X}_0), \quad \forall h \in \mathbb{Z}.$$

Properties.

1. $\Gamma(0) = \mathrm{cov}(\mathbf{X}_t),\ \forall t.$
2. $\Gamma(h) = \Gamma(-h)'$
3. In general $\Gamma(h)$ is not symmetric so that $\Gamma(h) \neq \Gamma(-h)$
4. $\Gamma(h)_{i,i}$ is autocovariance function of $(X_{t,i})_{t\in\mathbb{Z}}$

Correlation Matrix Function of Stationary T.S.

The correlation matrix function is given by

$$P(h) := \mathrm{corr}(X_h, X_0), \quad \forall h \in \mathbb{Z}.$$

In terms of the autocovariance matrix function we may write $P(h) = \Delta^{-1}\Gamma(h)\Delta^{-1}$, where $\Delta = \mathrm{diag}(\sqrt{\Gamma(0)_{11}}, \ldots, \sqrt{\Gamma(0)_{dd}})$.
(Recall univariate case)

$P(h)_{ij}$ and $P(h)_{ji}$ need not be the same. Often we have

$$\begin{aligned} P(h)_{ij} = \rho(X_{t+h,i}, X_{t,j}) &\approx 0, \quad \forall h > 0, \\ P(h)_{ji} = \rho(X_{t,i}, X_{t+h,j}) &\neq 0, \quad \text{some } h > 0, \end{aligned}$$

and series i is then said to lead series j.

The Cross Correlogram

Given time series data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ we calculate the sample covariance matrix function

$$\widehat{\Gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (\mathbf{X}_t - \overline{\mathbf{X}})(\mathbf{X}_{t+h} - \overline{\mathbf{X}})' \quad \text{where} \quad \overline{\mathbf{X}} = \sum_{t=1}^{n} \mathbf{X}_t / n.$$

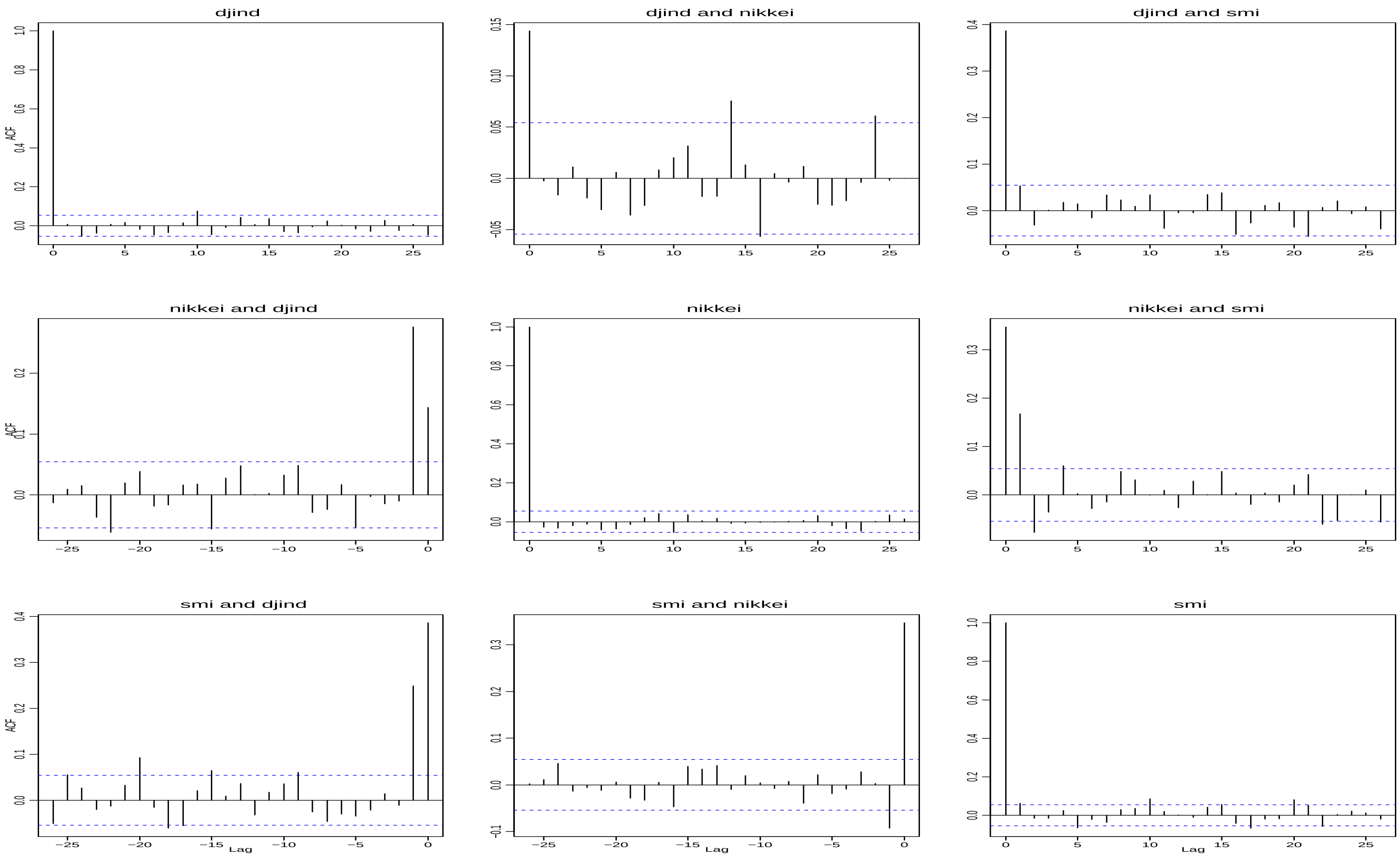
Hence we can construct sample correlation matrix function $\widehat{P}(h)$.

The cross correlogram is a matrix of plots. The (i, j) plot is $\{(h, \widehat{P}(h)_{ij}) : h = 0, 1, 2, \ldots\}$. (Recall univariate case)

Remarks.

1. Diagonal pictures are correlograms for individual series.
2. In S-Plus the (i, j) plot is $\{(-h, \widehat{P}(h)_{ij}) : h = 0, 1, 2, \ldots\}$ when $i > j$.

Cross Correlogram (DJ30, SMI, Nikkei Returns)



Multivariate White Noise Processes

A multivariate white noise process (WN) is a covariance stationary multivariate time series process whose correlation matrix function is given by

$$P(0) = P, \quad P(h) = 0, \; h \neq 0,$$

where P is any positive-definite correlation matrix.

A multivariate strict white noise process (SWN) is simply a process of iid random vectors.

(Recall univariate case)

G3. General Multivariate GARCH Model

Recall univariate GARCH-type model in (18).

The process $(\mathbf{X}_t)_{t\in\mathbb{Z}}$ is a multivariate GARCH model if it is strictly stationary and satisfies equations of the form

$$\mathbf{X}_t = \Sigma_t^{1/2}\mathbf{Z}_t \quad t \in \mathbb{Z}, \tag{27}$$

- $\Sigma_t^{1/2}$ is the Cholesky factor of a positive-definite matrix Σ_t which is measurable with respect to $\mathcal{F}_{t-1} = \sigma\{(\mathbf{X}_s)_{s\le t-1}\}$;
- $(\mathbf{Z}_t)_{t\in\mathbb{Z}}$ is multivariate SWN with mean zero and $\operatorname{cov}(\mathbf{Z}_t) = I_d$; (typically multivariate normal or (scaled) t innovations are used).

Conditional and Unconditional Moments

A covariance-stationary process satisfies

$$E(\mathbf{X}_t \mid \mathcal{F}_{t-1}) = E(\Sigma_t^{1/2}\mathbf{Z}_t \mid \mathcal{F}_{t-1}) = \Sigma_t^{1/2}E(\mathbf{Z}_t) = \mathbf{0},$$

(martingale difference property) so process is a multivariate white noise process. Moreover Σ_t is the conditional covariance matrix since

$$\mathrm{cov}(\mathbf{X}_t \mid \mathcal{F}_{t-1}) = E(\mathbf{X}_t\mathbf{X}_t' \mid \mathcal{F}_{t-1}) = \Sigma_t^{1/2}E(\mathbf{Z}_t\mathbf{Z}_t')\Sigma_t^{1/2\,\prime} = \Sigma_t.$$

The unconditional moments are $E(\mathbf{X}_t) = \mathbf{0}$ and

$$\Sigma = \mathrm{cov}(\mathbf{X}_t) = E\left(\mathrm{cov}(\mathbf{X}_t \mid \mathcal{F}_{t-1})\right) + \mathrm{cov}(E(\mathbf{X}_t \mid \mathcal{F}_{t-1})) = E(\Sigma_t).$$

Conditional Correlation

We have the decomposition

$$\Sigma_t = \Delta_t P_t \Delta_t, \quad \Delta_t = \text{diag}(\sigma_{t,1}, \dots, \sigma_{t,d}),$$

where the diagonal matrix Δ_t is the volatility matrix and contains the volatilities for the component series $(X_{t,k})_{t \in \mathbb{Z}}$, $k = 1, \dots, d$. P_t is the conditional correlation matrix and has elements given by

$$P_{t,ij} = \Sigma_{t,ij} \big/ (\sigma_{t,i} \sigma_{t,j}).$$

To build parametric models we parameterize the dependence of Σ_t (or of Δ_t and P_t) on the past values of the process in such a way that Σ_t always remains symmetric and positive definite.

Multivariate ARMA-GARCH Models

As before (recall univariate case) we can consider general models of the form

$$\mathbf{X}_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2}\mathbf{Z}_t$$

where $\boldsymbol{\mu}_t$ is a $\mathcal{F}_{t-1}$-measurable conditional mean vector.

We might choose to parametrise it using a vector autoregressive (VAR) or vector ARMA (VARMA) model.

Example: VAR(1)

$$\boldsymbol{\mu}_t = \boldsymbol{\mu} + \Phi\mathbf{X}_{t-1}$$

where all eigenvalues of Φ are less than one in absolute value.

G4. The Constant Conditional Correlation Model

A useful model is the CCC-GARCH model, proposed by Bollerslev (1990), where it is assumed that

- Conditional correlation is constant over time: $P_t \equiv P$ for all t
- Volatilities follow standard univariate GARCH models.
  For example, in a CCC-GARCH(1,1) model

$$\sigma_{t,i}^2 = \alpha_0^i + \alpha_1^i X_{t-1,i}^2 + \beta^i \sigma_{t-1,i}^2.$$

- The $\mathbf{Z}_t$ are SWN$(0, \mathsf{I}_d)$ with normal or scaled t distribution.

Fitting the CCC-model

- Full maximum likelihood estimation, using the sample correlation matrix (or robust alternative) as starting value.

- Estimation of P by sample correlation matrix, and estimation of remaining parameters by ML.

- Simple two-stage procedure relying on the observation that in CCC model $\mathbf{X}_t = \Delta_t \mathbf{Y}_t$ where the sequence $(\mathbf{Y}_t)_{t \in \mathbb{Z}}$ is iid with covariance matrix P .

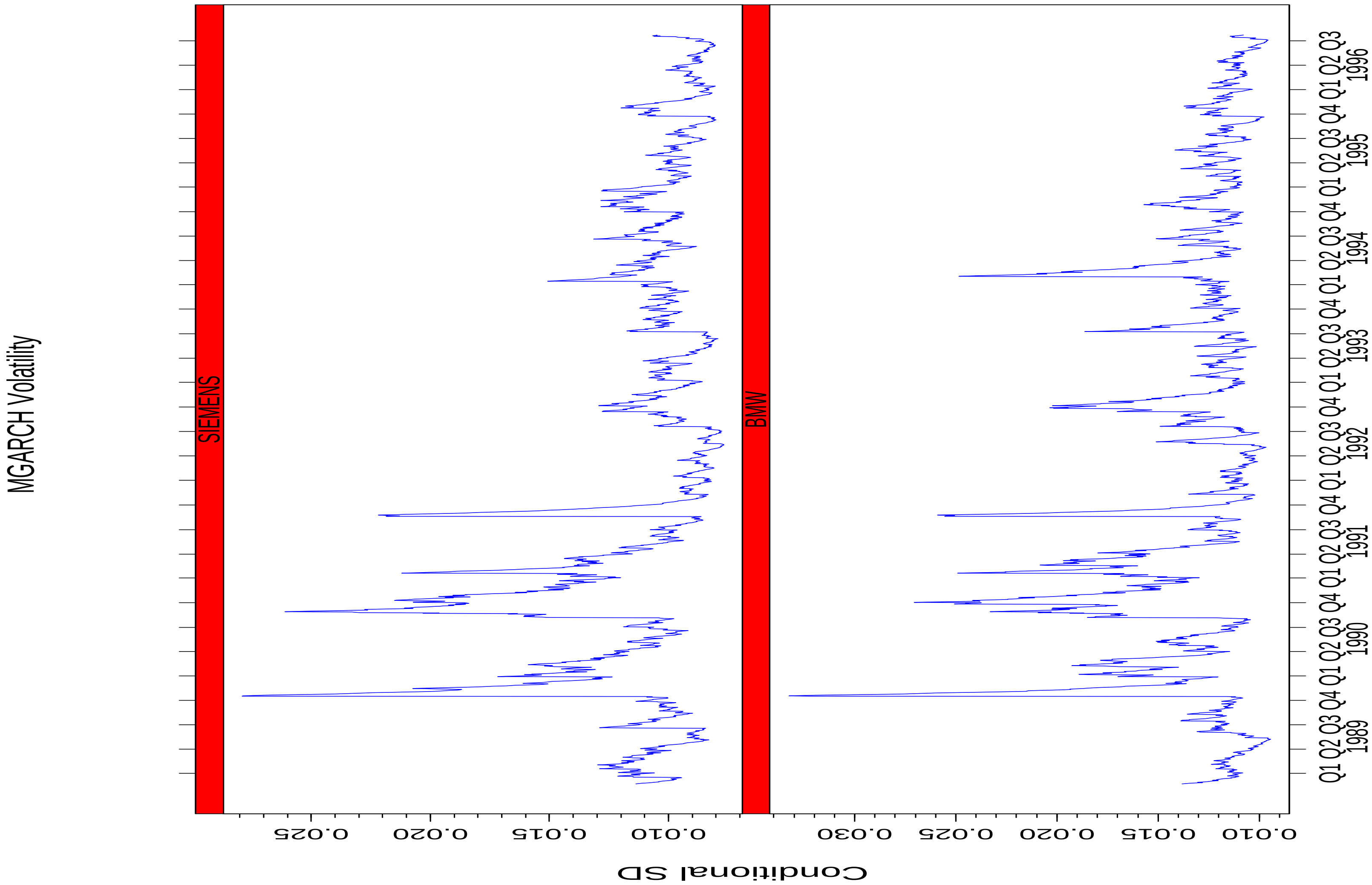
  Stage 1. Fit univariate GARCH models to each component to determine dynamics of Δ_t. Form residuals $\widehat{\mathbf{Y}}_t := \widehat{\Delta}_t^{-1} \mathbf{X}_t$.

  Stage 2. Estimate distribution of $\mathbf{Y}_t$, either by fitting parametric model (e.g. t-distribution) to residuals or by "historical simulation" using residuals.

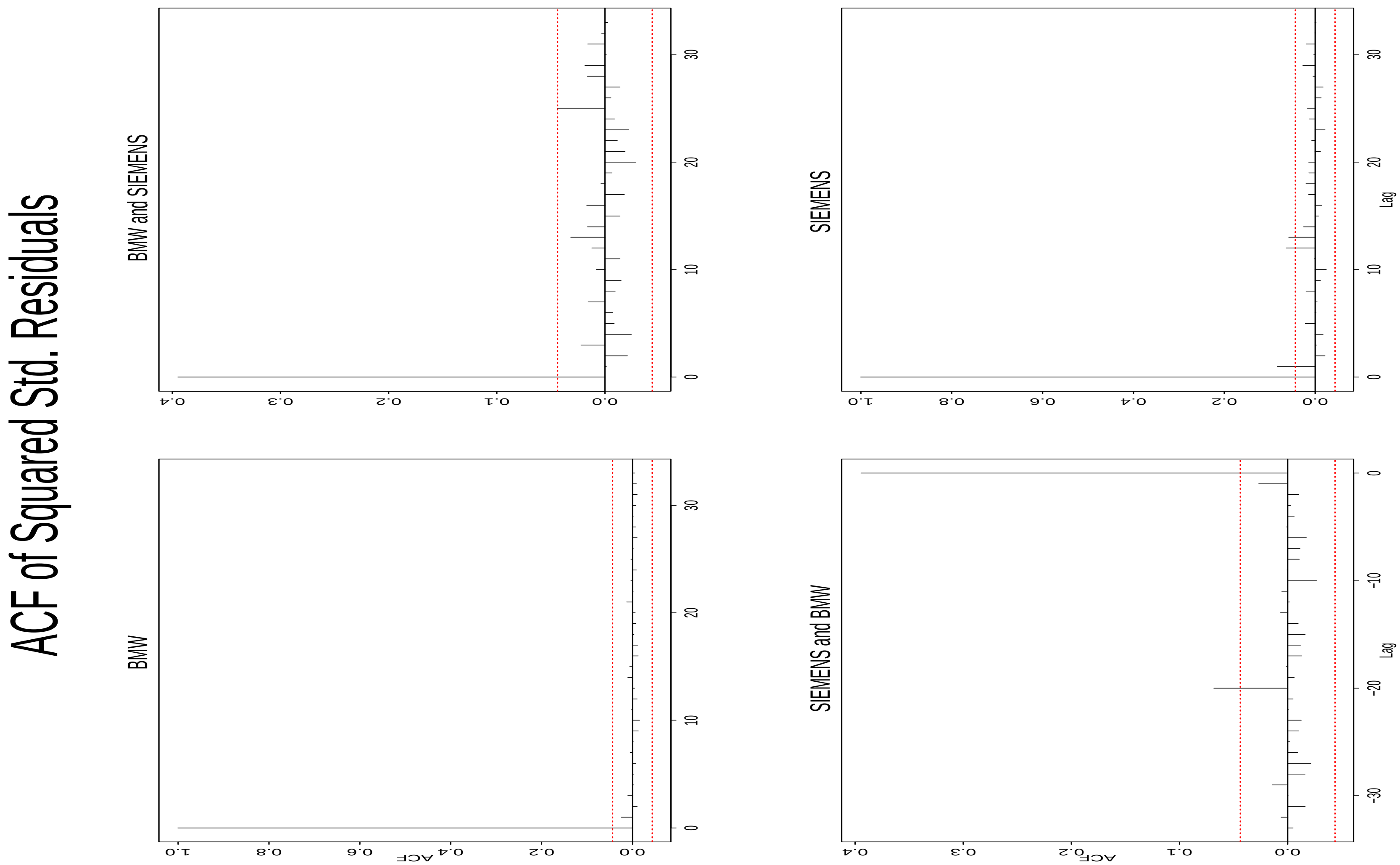
CCC-GARCH Analysis of BMW-Siemens Data

```
Mean Equation: dataforgarch ~ 1
Conditional Variance Equation:  ~ ccc(1, 1)
Conditional Distribution:  t with estimated parameter 4.671218 and standard error 0.3307812

Estimated Coefficients:
--------------------------------------------------------------
                   Value Std.Error t value  Pr(>|t|)
          C(1) 1.425e-04 2.413e-04  0.5907 2.774e-01
          C(2) 2.158e-04 1.964e-04  1.0990 1.359e-01
       A(1, 1) 6.904e-06 1.802e-06  3.8324 6.542e-05
       A(2, 2) 3.492e-06 1.013e-06  3.4459 2.904e-04
 ARCH(1; 1, 1) 4.597e-02 8.887e-03  5.1727 1.269e-07
 ARCH(1; 2, 2) 4.577e-02 1.001e-02  4.5714 2.570e-06
GARCH(1; 1, 1) 9.121e-01 1.667e-02 54.7174 0.000e+00
GARCH(1; 2, 2) 9.264e-01 1.515e-02 61.1655 0.000e+00

Estimated Conditional Constant Correlation Matrix:
--------------------------------------------------------------
           BMW SIEMENS
    BMW 1.0000  0.6818
SIEMENS 0.6818  1.0000
```

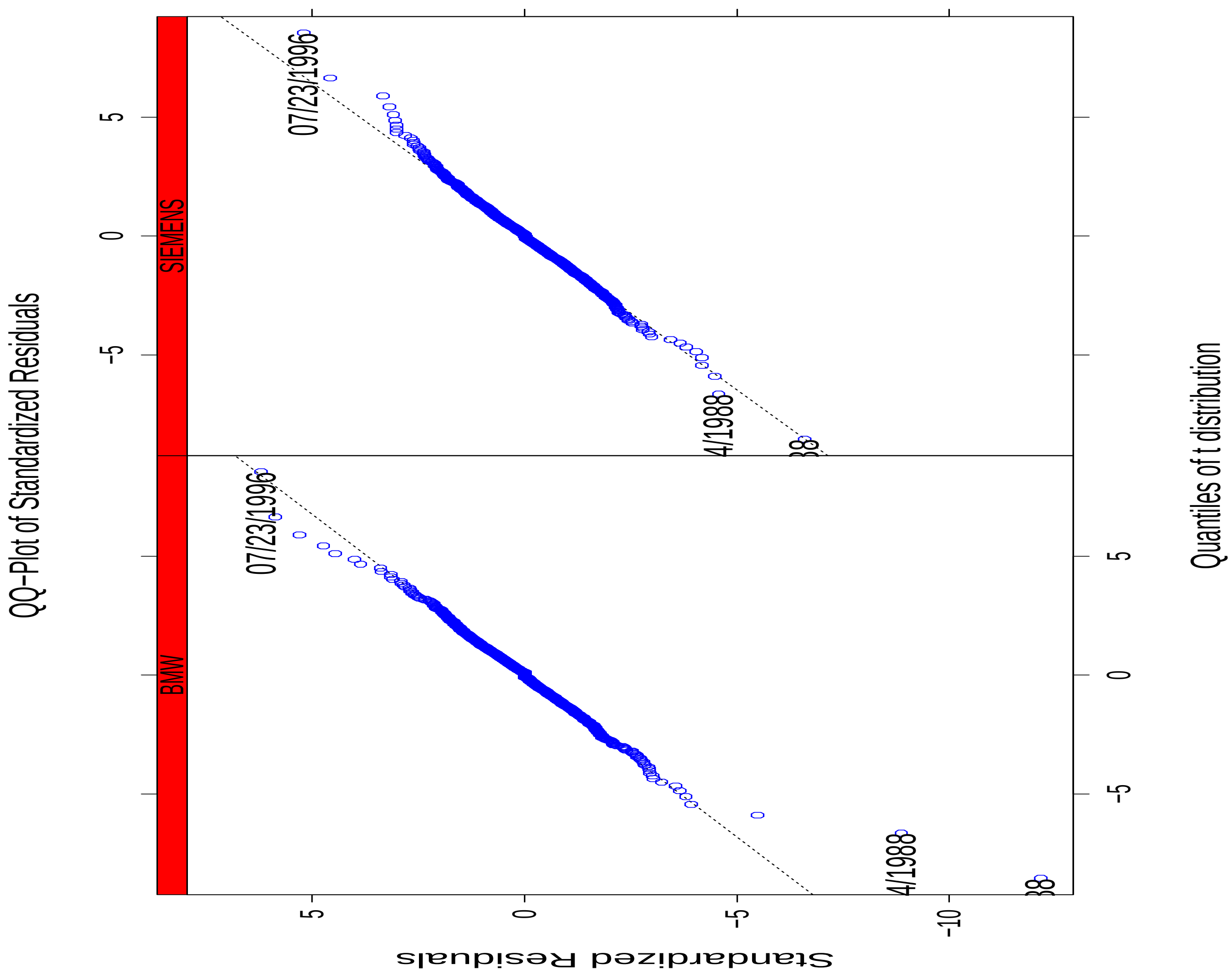
CCC-GARCH(1,1) Analysis



CCC-GARCH(1,1) Analysis II



CCC-GARCH(1,1) Analysis III



G5. The DVEC Family

There have been many proposals for models that allow dynamically changing correlations; we choose one of the more successful ones.

In a DVEC model the conditional covariance matrix satisfies

$$\Sigma_t = A_0 + \sum_{i=1}^{p} A_i \circ \mathbf{X}_{t-i}\mathbf{X}'_{t-i} + \sum_{j=1}^{q} B_j \circ \Sigma_{t-j},$$

where $A_0, A_i, B_j \in \mathbb{R}^{d\times d}$ are symmetric parameter matrices.
Compare with (19).

Remarks

1. The symbol $\circ$ denotes componentwise multiplication of matrices.
2. First order models generally good enough.

Guaranteeing Positive Definiteness

To ensure positive definiteness of Σ_t we can require all parameter matrices to be p.d. although not all software does this.

A parameterisation that ensures positive definiteness sets

$$A_i = A_i^{1/2} A_i^{1/2\,\prime},\ i = 0, 1, \ldots, p, \quad B_j = B_j^{1/2} B_j^{1/2\,\prime}, j = 1, \ldots, q,$$

where the $A_i^{1/2}$ and $B_j^{1/2}$ are lower-triangular matrices.

A Simpler Model (DVEC.vec.vec)

The number of parameters can be further reduced by setting

$$A_0 = A_0^{1/2} A_0^{1/2\,\prime},\ A_i = \mathbf{a}_i \mathbf{a}_i',\ i = 1, \ldots, p, \quad B_j = \mathbf{b}_j \mathbf{b}_j',\ j = 1, \ldots, q,$$

where $A_0^{1/2\,\prime}$ is lower triangular and the $\mathbf{a}_i$ and $\mathbf{b}_j$ are vectors.

G6. Risk Management with Multivariate GARCH

For simplicity consider a portfolio of several stocks which has value V_t at time t. In terms of the log returns for the next time period the loss follows (5) and is

$$L_{t+1} = -V_t \sum_{i=1}^{d} \omega_{t,i} \left(e^{X_{t+1,i}} - 1\right) \approx -V_t \sum_{i=1}^{d} \omega_{t,i} X_{t+1,i} = L_{t+1}^{\Delta},$$

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

We assume the time series $(\mathbf{X}_t)$ follows a model with structure

$$X_t = \boldsymbol{\mu}_t + \Sigma_t^{1/2} \mathbf{Z}_t, \quad \boldsymbol{\mu}_t, \Sigma_t \in \mathcal{F}_{t-1},$$

where $(\mathbf{Z}_t)$ are SWN innovations with mean zero and covariance matrix I_d, such as a multivariate GARCH model.

Normal Variance Mixture Innovations

If the innovations come from a normal or normal variance mixture distribution, such as scaled Student t, the conditional distribution of linearized loss L^{Δ}_{t+1} given $\mathcal{F}_t$ will be of the same type.

- If $Z_{t+1} \sim N_d(\mathbf{0}, \mathsf{I}_d)$ then

$$L^{\Delta}_{t+1} \mid \mathcal{F}_t \sim N\left(-V_t \boldsymbol{\omega}'_t \boldsymbol{\mu}_{t+1}, V_t^2 \boldsymbol{\omega}'_t \Sigma_{t+1} \boldsymbol{\omega}_t\right).$$

- If $Z_{t+1} \sim t_d(\nu, \mathbf{0}, (\nu-2)\mathsf{I}_d/\nu)$ then

$$L^{\Delta}_{t+1} \mid \mathcal{F}_t \sim t_1\left(\nu, -V_t \boldsymbol{\omega}'_t \boldsymbol{\mu}_{t+1}, \frac{\nu-2}{\nu} V_t^2 \boldsymbol{\omega}'_t \Sigma_{t+1} \boldsymbol{\omega}_t\right).$$

(See definition of multivariate t)

Risk Measures

Conditional risk measures take form

$$\begin{aligned}
\mathrm{VaR}_{\alpha}^{t} &= -V_t \boldsymbol{\omega}_t' \boldsymbol{\mu}_{t+1} + V_t \sqrt{\boldsymbol{\omega}_t' \Sigma_{t+1} \boldsymbol{\omega}_t}\, q_\alpha(F_Z) \\
\mathrm{ES}_{\alpha}^{t} &= -V_t \boldsymbol{\omega}_t' \boldsymbol{\mu}_{t+1} + V_t \sqrt{\boldsymbol{\omega}_t' \Sigma_{t+1} \boldsymbol{\omega}_t}\, E(Z \mid Z > q_\alpha(F_Z)).
\end{aligned}$$

where Z denotes variate with univariate standard normal or scaled t distribution

As before, we estimate $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1} by GARCH-based predictions. For scaled t we require (25) and (26) and ν is estimated in GARCH model.

H. 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

H1. 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 $F^{-1}(U) \stackrel{\mathrm{d}}{=} X$ 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

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

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 $\text{Ran}F_1 \times \text{Ran}F_2 \ldots \times \text{Ran}F_d$.

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.

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

where Φ is df of standard normal.

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

H2. 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).

H3. 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

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

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 $\left(\left(1 - \frac{\log U_1}{X}\right)^{-1/\beta}, \ldots, \left(1 - \frac{\log U_d}{X}\right)^{-1/\beta}\right)$.

H4. 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

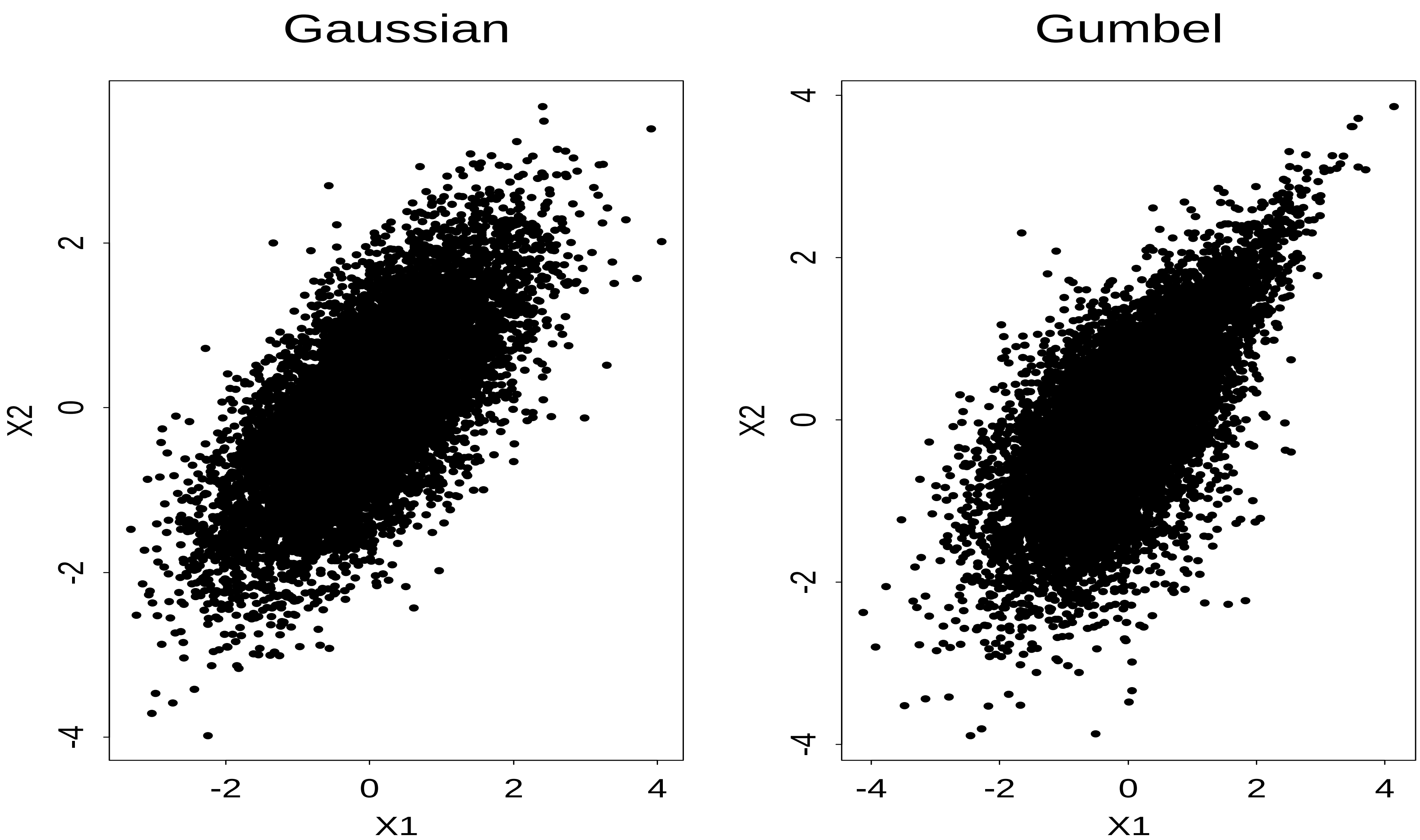
Consider the random vector $(X_1, X_2)'$.

"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%.

H5. 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

$$\rho_\tau(X_1, X_2) = \frac{2}{\pi} \arcsin\left(\rho(X_1, X_2)\right). \tag{28}$$

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

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

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

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

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: $(U_1, U_2) \stackrel{\mathsf{d}}{=} (1-U_1, 1-U_2)$.

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 $\beta > 1$.

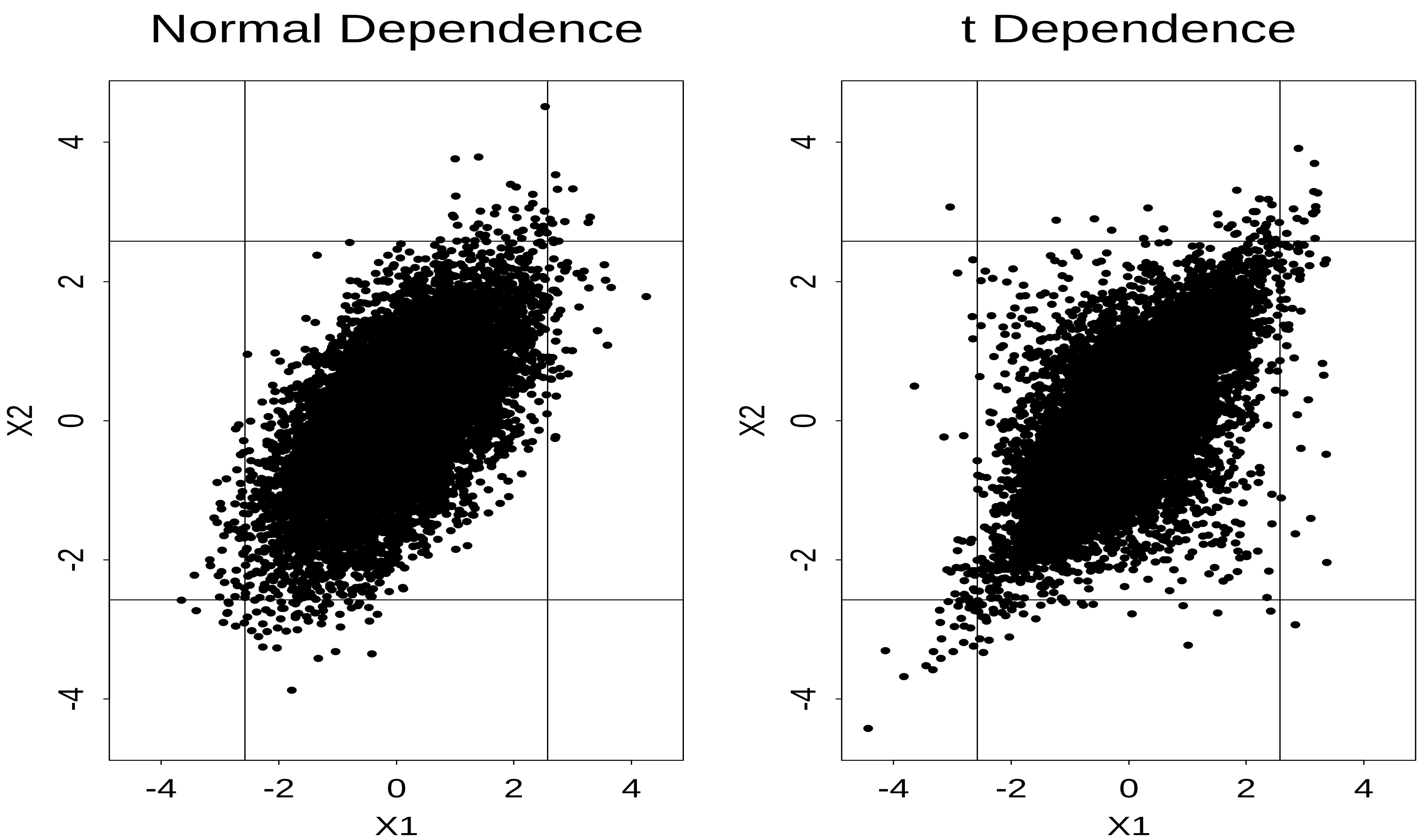
$$\lambda_u = 2 - 2^{1/\beta}.$$

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.

H6. 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

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

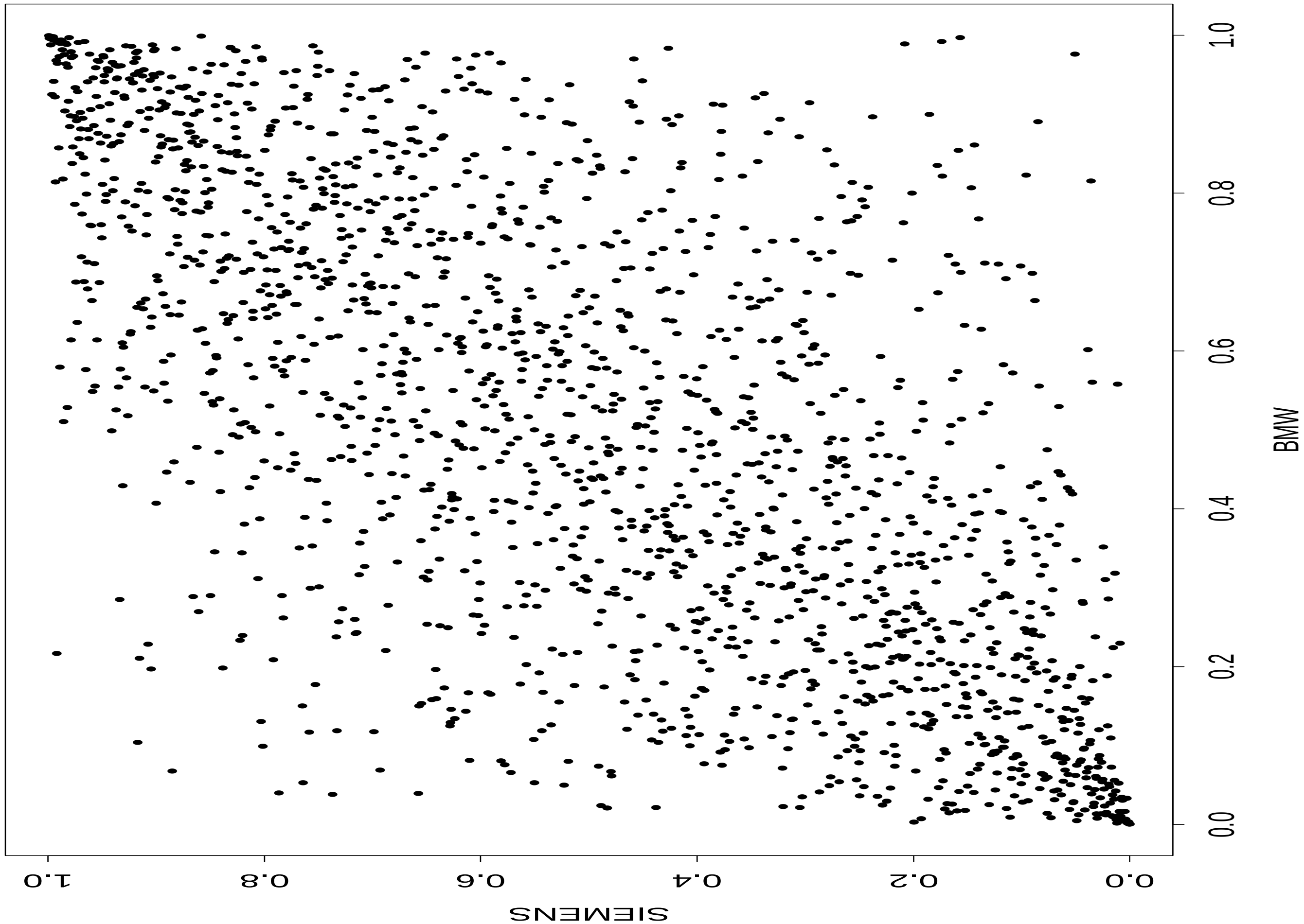
and fit parametric copula C by maximum likelihood.

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.

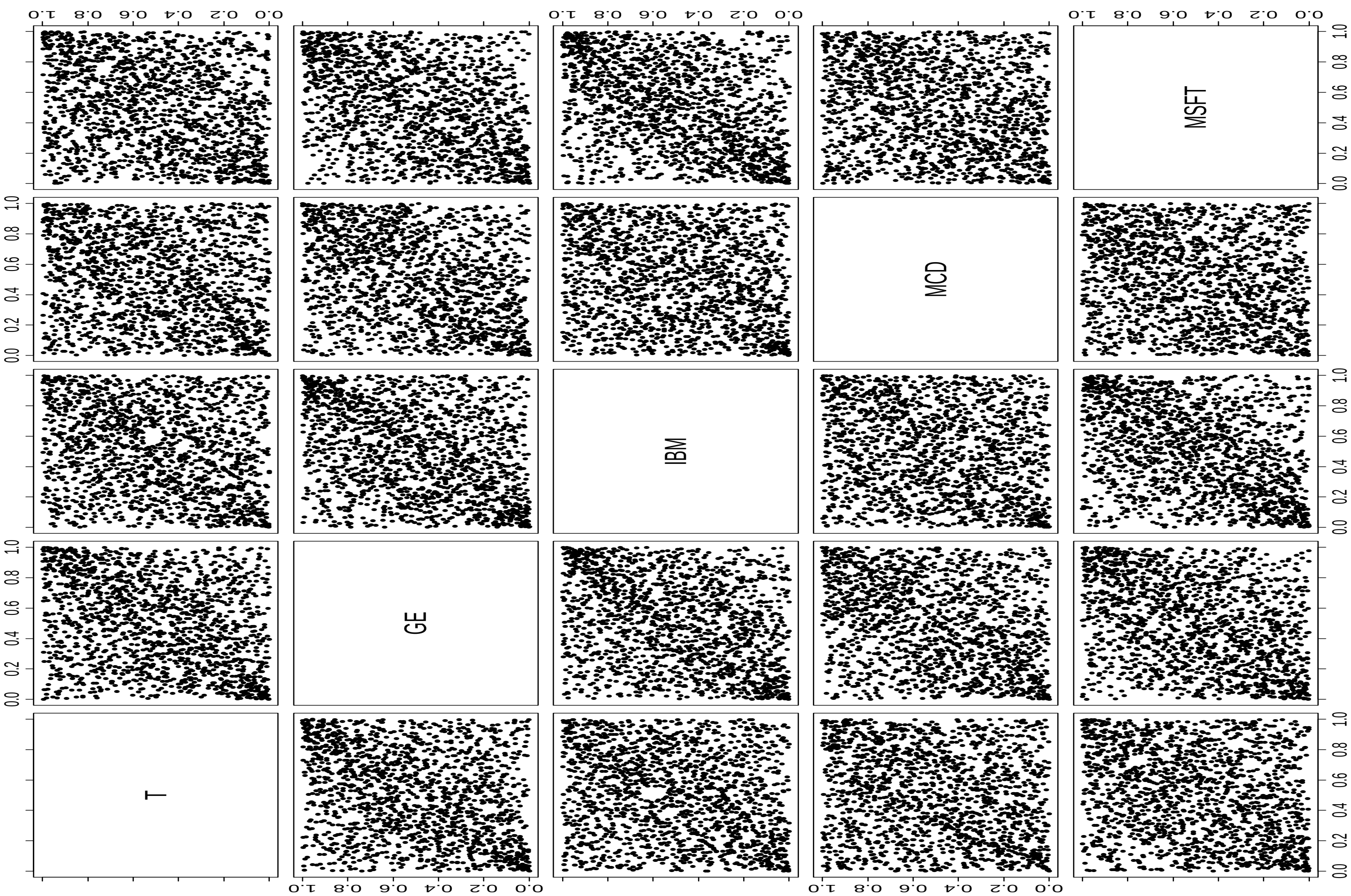
Remark. Formal methods for goodness-of-fit also available.

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:

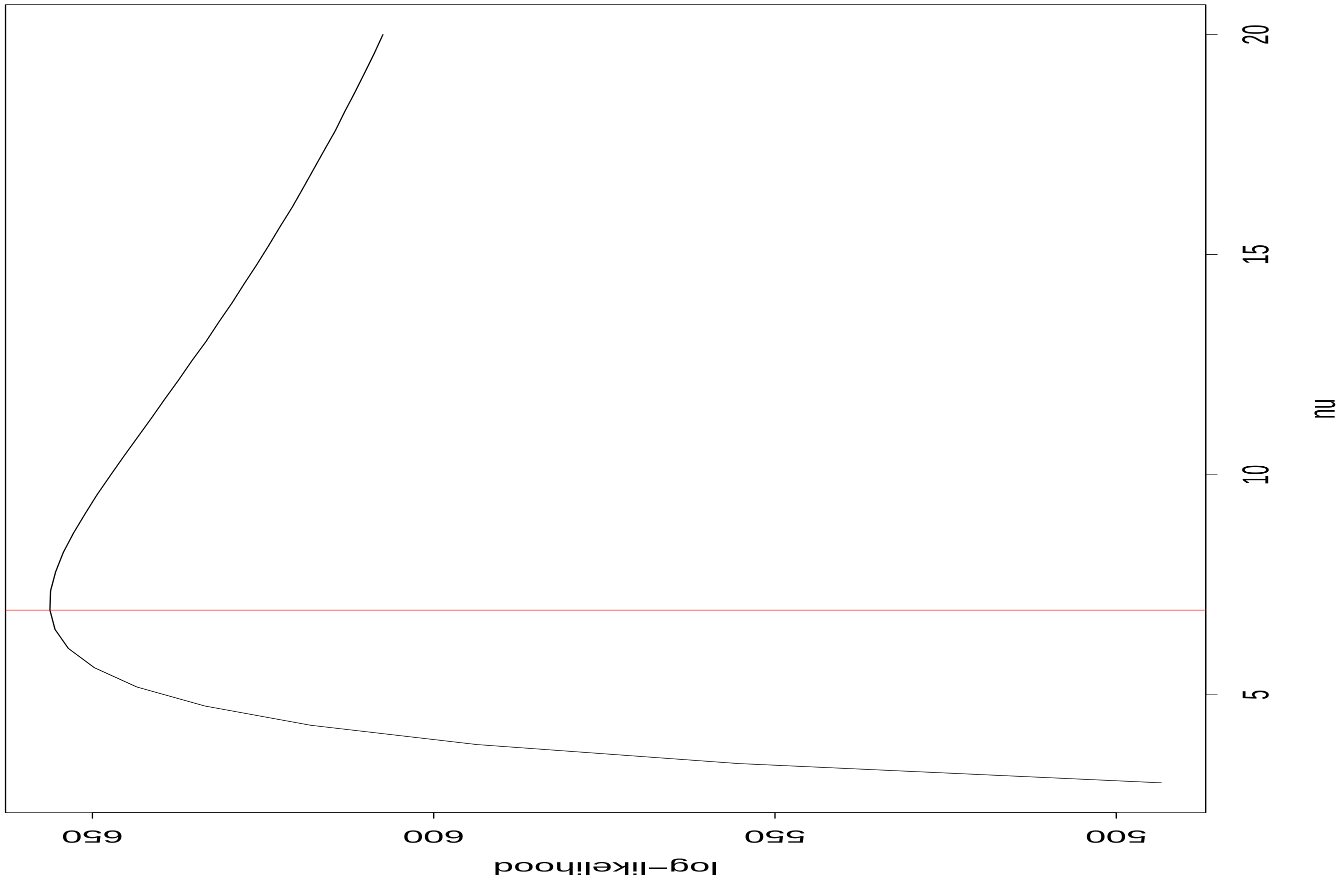
- Estimate Kendall's tau matrix from the data.
- Recall that if $\mathbf{X}$ is meta-t with df $C^t_{\nu,P}(F_1,\ldots,F_d)$ then $\rho_\tau(X_i,X_j) = \frac{2}{\pi}\arcsin(P_{i,j})$. Follows from (28).
- Estimate $\widehat{P}_{i,j} = \sin\left(\frac{\pi}{2}\widehat{\rho}_\tau(X_i,X_j)\right)$. Check 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.

I. Maxima and Worst Cases

1. Limiting Behaviour of Sums and Maxima

2. Extreme Value Distributions

3. The Fisher–Tippett Theorem

4. The Block Maxima Method

5. S&P Example

I1. Limiting Behaviour of Maxima

Let $X_1, X_2, \ldots$ be iid random variables with distribution function (df) F. In risk management applications these could represent financial losses, operational losses or insurance losses.

Let $M_n = \max(X_1, \ldots, X_n)$ be worst–case loss in a sample of n losses. Clearly

$$P(M_n \le x) = P(X_1 \le x, \ldots, X_n \le x) = F^n(x)\,.$$

It can be shown that, almost surely, $M_n \stackrel{n\to\infty}{\longrightarrow} x_F$, where $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} \le \infty$ is the right endpoint of F.

But what about normalized maxima?

Limiting Behaviour of Sums or Averages

(See [Embrechts et al., 1997], Chapter 2.)

We are familiar with the central limit theorem.

Let $X_1, X_2, \ldots$ be iid with finite mean μ and finite variance σ^2. Let $S_n = X_1 + X_2 + \ldots + X_n$. Then

$$P\left((S_n - n\mu)\,/\sqrt{n\sigma^2} \le x\right) \overset{n\to\infty}{\longrightarrow} \Phi(x)\,,$$

where Φ is the distribution function of the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du\,.$$

Note, more generally, the limiting distributions for appropriately normalized sample sums are the class of α–stable distributions; Gaussian distribution is a special case.

Limiting Behaviour of Sample Extrema

(See [Embrechts et al., 1997], Chapter 3.)

Let $X_1, X_2, \ldots$ be iid from F and let $M_n = \max(X_1, \ldots, X_n)$.

Suppose we can find sequences of real numbers $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$, the sequence of normalized maxima, converges in distribution, i.e.

$$P\left((M_n - b_n)/a_n \leq x\right) = F^n(a_n x + b_n) \stackrel{n\to\infty}{\longrightarrow} H(x),$$

for some non–degenerate df $H(x)$.

If this condition holds we say that F is in the maximum domain of attraction of H, abbreviated $F \in \text{MDA}(H)$. Note that such an H is determined up to location and scale, i.e. will specify a unique type of distribution.

I2. Generalized Extreme Value Distribution

The GEV has df

$$H_\xi(x) = \begin{cases} \exp\left(-(1+\xi x)^{-1/\xi}\right) & \xi \neq 0\,, \\ \exp\left(-e^{-x}\right) & \xi = 0, \end{cases}$$

where $1 + \xi x > 0$ and ξ is the shape parameter. Note, this parametrization is continuous in ξ. For

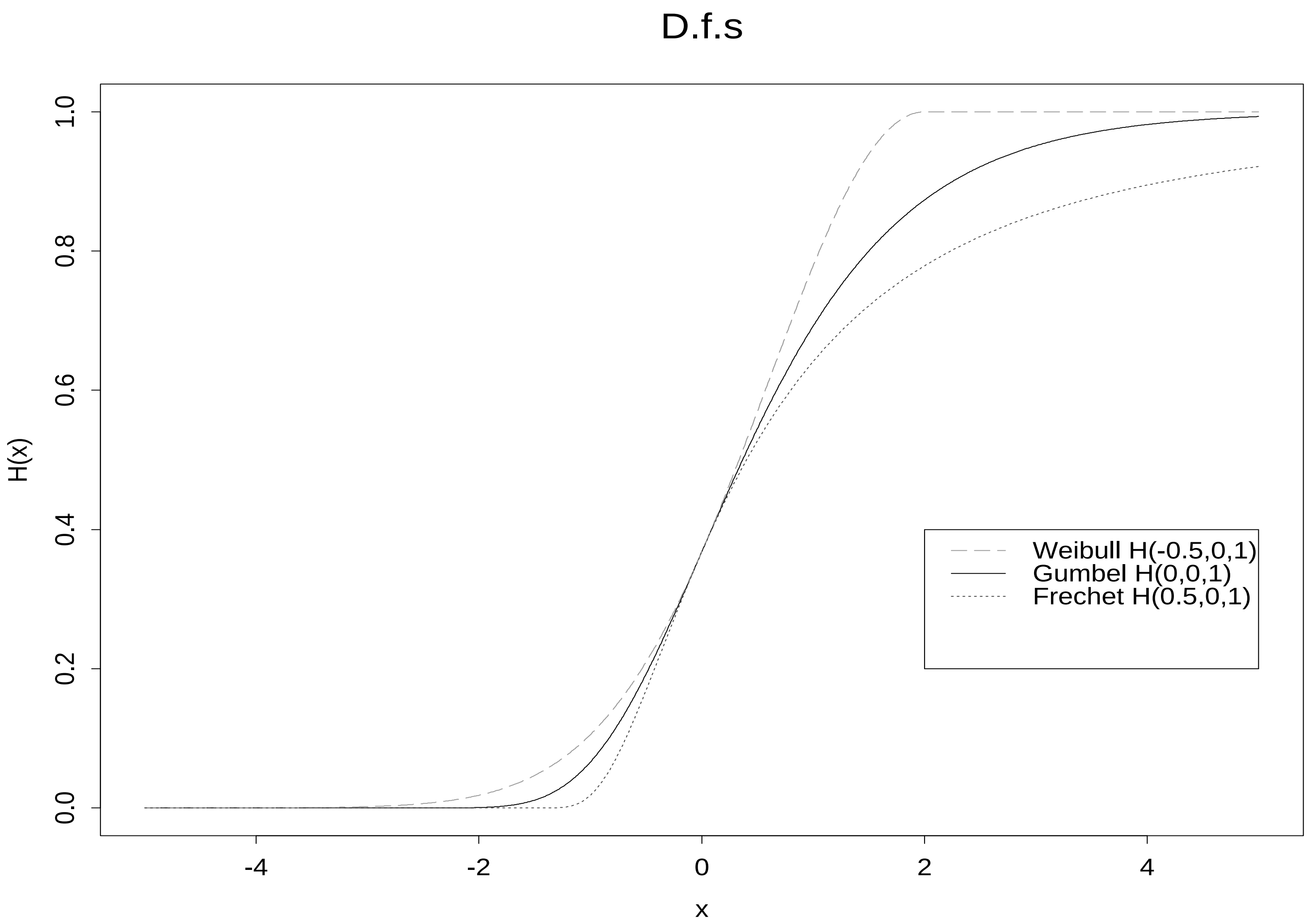
$\xi > 0$ $\quad H_\xi$ is equal in type to classical Fréchet df

$\xi = 0$ $\quad H_\xi$ is equal in type to classical Gumbel df

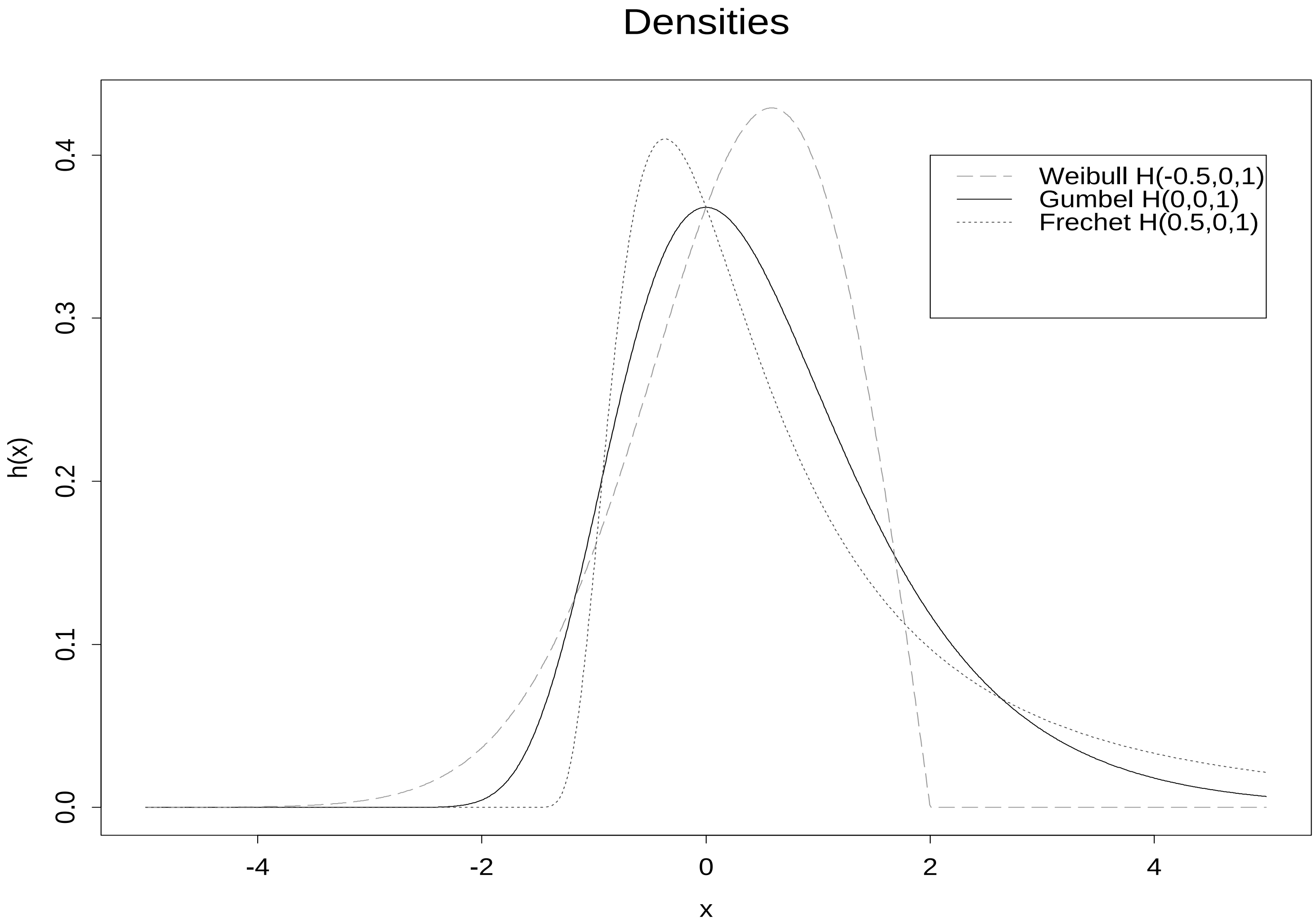
$\xi < 0$ $\quad H_\xi$ is equal in type to classical Weibull df.

We introduce location and scale parameters μ and $\sigma > 0$ and work with $H_{\xi,\mu,\sigma}(x) := H_\xi((x-\mu)/\sigma)$. Clearly $H_{\xi,\mu,\sigma}$ is of type H_ξ.

GEV: distribution functions for various ξ



GEV: densities for various ξ



I3. Fisher–Tippett Theorem (1928)

Theorem: If $F \in \text{MDA}(H)$ then H is of the type H_ξ for some ξ.

"If suitably normalized maxima converge in distribution to a non–degenerate limit, then the limit distribution must be an extreme value distribution."

Remark 1: Essentially all commonly encountered continuous distributions are in the maximum domain of attraction of an extreme value distribution.

Remark 2: We can always choose normalizing sequences a_n and b_n so that the limit law H_ξ appears in standard form (without relocation or rescaling).

Fisher-Tippett: Examples

Recall: $F \in \mathrm{MDA}(H_\xi)$, iff there are sequences a_n and b_n with

$$P\left(\left(M_n - b_n\right)/a_n \leq x\right) = F^n\left(a_n x + b_n\right) \stackrel{n\to\infty}{\longrightarrow} H(x).$$

We have the following examples:

- The exponential distribution, $F(x) = 1 - e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$, is in MDA(H_0) (Gumbel-case). Take $a_n = 1/\lambda$, $b_n = (\log n)/\lambda$.

- The Pareto distribution,

$$F(x) = 1 - \left(\frac{\kappa}{\kappa + x}\right)^\alpha, \quad \alpha, \kappa > 0, \quad x \geq 0,$$

is in MDA($H_{1/\alpha}$) (Fréchet case). Take $a_n = \kappa n^{1/\alpha}/\alpha$, $b_n = \kappa n^{1/\alpha} - \kappa$.

I4. Using Fisher–Tippett: Block Maxima Method

Assume that we have a large enough block of n iid random variables so that the limit result is more or less exact, i.e. $\exists a_n > 0$, $b_n \in \mathbb{R}$ such that, for some ξ,

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \approx H_\xi(x)\,.$$

Now set $y = a_n x + b_n$. $\quad P\left(M_n \leq y\right) \approx H_\xi\left(\frac{y-b_n}{a_n}\right) = H_{\xi,b_n,a_n}(y)$.
We wish to estimate ξ, b_n and a_n.

Implication: We collect data on block maxima and fit the three–parameter form of the GEV. For this we require a lot of raw data so that we can form sufficiently many, sufficiently large blocks.

ML Inference for Maxima

We have block maxima data $\mathbf{y} = \left(M_n^{(1)}, \ldots, M_n^{(m)}\right)'$ from m blocks of size n. We wish to estimate $\boldsymbol{\theta} = (\xi, \mu, \sigma)'$. We construct a log–likelihood by assuming we have independent observations from a GEV with density $h_{\boldsymbol{\theta}}$,

$$l(\boldsymbol{\theta}; \mathbf{y}) = \log \left(\prod_{i=1}^{m} h_{\boldsymbol{\theta}} \left(M_n^{(i)} \right) 1_{\left\{1+\xi\left(M_n^{(i)}-\mu\right)/\sigma>0\right\}} \right),$$

and maximize this w.r.t. $\boldsymbol{\theta}$ to obtain the MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\xi}, \widehat{\mu}, \widehat{\sigma})'$.

Clearly, in defining blocks, bias and variance must be traded off. We reduce bias by increasing the block size n; we reduce variance by increasing the number of blocks m.

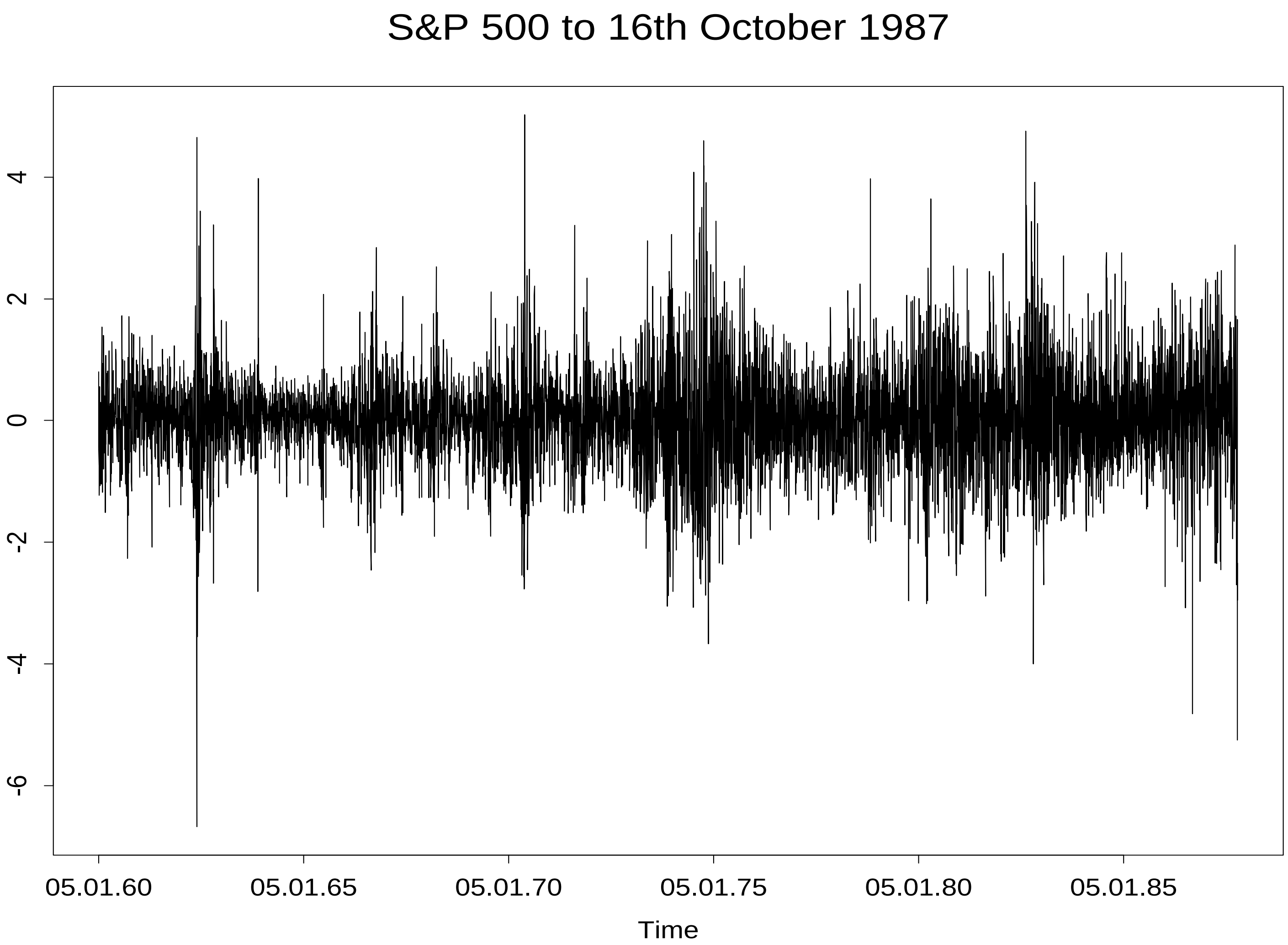
I5. An Example: S&P 500

It is the early evening of Friday the 16th October 1987. In the equity markets it has been an unusually turbulent week, which has seen the S&P 500 index fall by 9.21%. On that Friday alone the index is down 5.25% on the previous day, the largest one–day fall since 1962. At our disposal are all daily closing values of the index since 1960.

We analyse annual maxima of daily percentage falls in the index. These values $M_{260}^{(1)}, \ldots, M_{260}^{(28)}$ are assumed to be iid from $H_{\xi,\mu,\sigma}$.

Remark. Although we have only justified this choice of limiting distribution for maxima of iid data, it turns out that the GEV is also the correct limit for maxima of stationary time series, under some technical conditions on the nature of the dependence. These conditions are fulfilled, for example, by GARCH processes.

S&P 500 Return Data



Assessing the Risk in S&P

We will address the following two questions:

- What is the probability that next year's maximum exceeds all previous levels?

- What is the 40–year return level $R_{260,40}$?

In the first question we assess the probability of observing a new record. In the second problem we define and estimate a rare stress or scenario loss.

Return Levels

$R_{n,k}$, the k n–block return level, is defined by

$$P\left(M_n > R_{n,k}\right) = 1/k\,;$$

i.e. it is that level which is exceeded in one out of every k n–blocks, on average.

We use the approximation

$$R_{n,k} \approx H^{-1}_{\xi,\mu,\sigma}(1-1/k) \approx \mu + \sigma\left((-\log(1-1/k))^{-\xi} - 1\right)/\xi\,.$$

We wish to estimate this functional of the unknown parameters of our GEV model for maxima of n–blocks.

S–Plus Maxima Analysis with EVIS

```
> out <- gev(-sp,"year")
> out
$n.all: [1] 6985

$n: [1] 28

$data:
     1960     1961     1962     1963     1964     1965     1966     1967
 2.268191 2.083017 6.675635 2.806479 1.253012 1.757765 2.460411 1.558183
     1968     1969     1970     1971     1972     1973     1974     1975
 1.899367 1.903001 2.768166 1.522388 1.319013 3.051598 3.671256 2.362394
     1976     1977     1978     1979     1980     1981     1982     1983
 1.797353 1.625611 2.009257 2.957772 3.006734 2.886327 3.996544 2.697254
     1984     1985     1986     1987
 1.820587 1.455301 4.816644 5.253623

$par.ests:
        xi     sigma       mu
 0.3343843 0.6715922 1.974976

$par.ses:
     xi    sigma        mu
 0.2081 0.130821 0.1512828

$nllh.final:
[1] 38.33949
```

S&P Example (continued)

Answers:

- Probability is estimated by

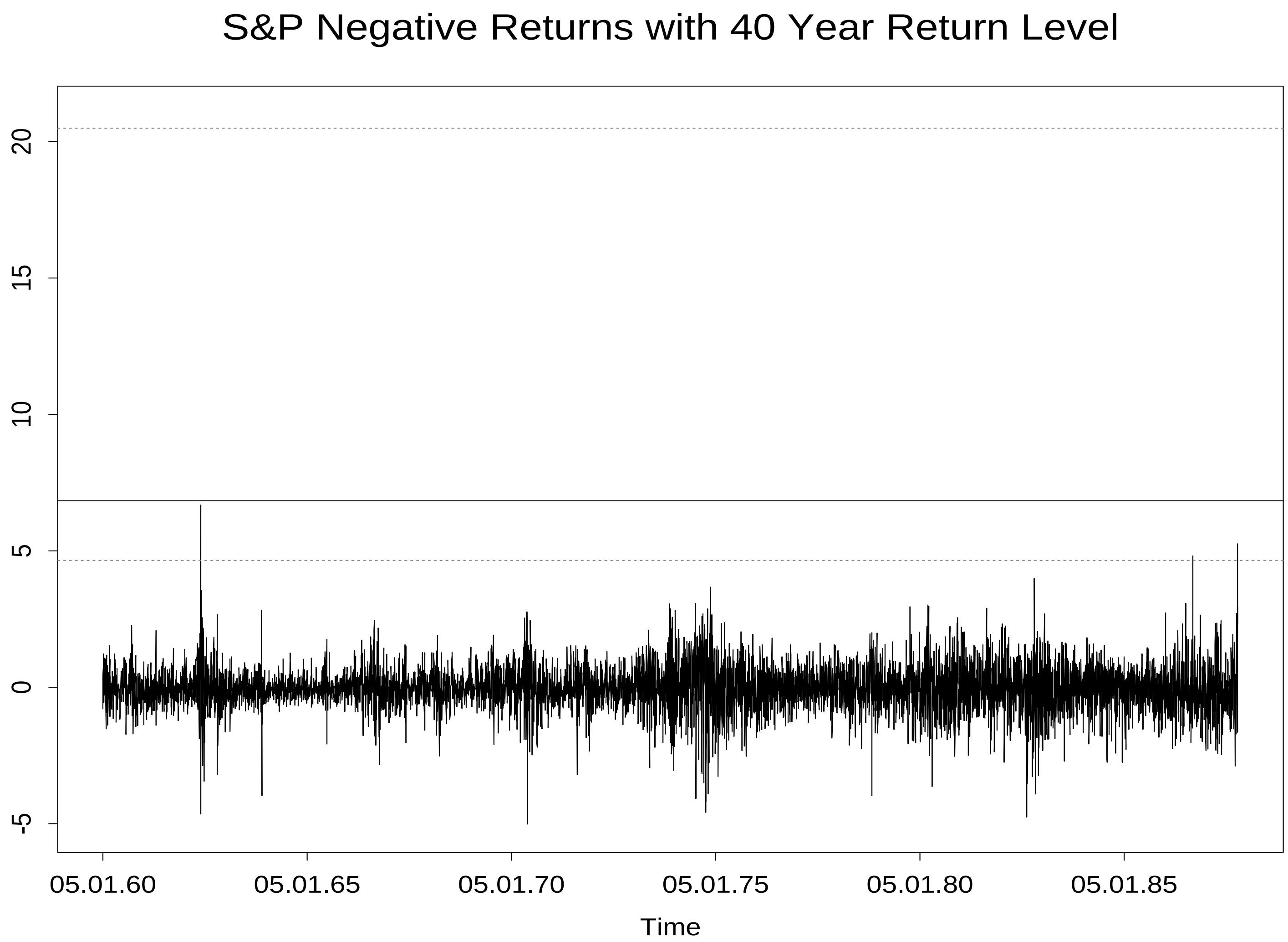
$$1 - H_{\hat{\xi},\hat{\mu},\hat{\sigma}}\left(\max\left(M_{260}^{(1)}, \ldots, M_{260}^{(28)}\right)\right) = 0.027\,.$$

- $R_{260,40}$ is estimated by

$$H_{\hat{\xi},\hat{\mu},\hat{\sigma}}^{-1}(1 - 1/40) = 6.83\,.$$

It is important to construct confidence intervals for such statistics. We use asymptotic likelihood ratio ideas to construct asymmetric intervals – the so–called profile likelihood method.

Estimated 40–Year Return Level



J. The Peaks–over–Thresholds (POT) Method

1. The Generalized Pareto Distribution (GPD)
2. The POT Method: Theoretical Foundations
3. Modelling Tails and Quantiles of Distributions
4. The Danish Fire Loss Analysis
5. Expected Shortfall and Mean Excess Plot

J1. Generalized Pareto Distribution

The GPD is a two parameter distribution with df

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \xi \neq 0\,, \\ 1 - \exp(-x/\beta) & \xi = 0\,, \end{cases}$$

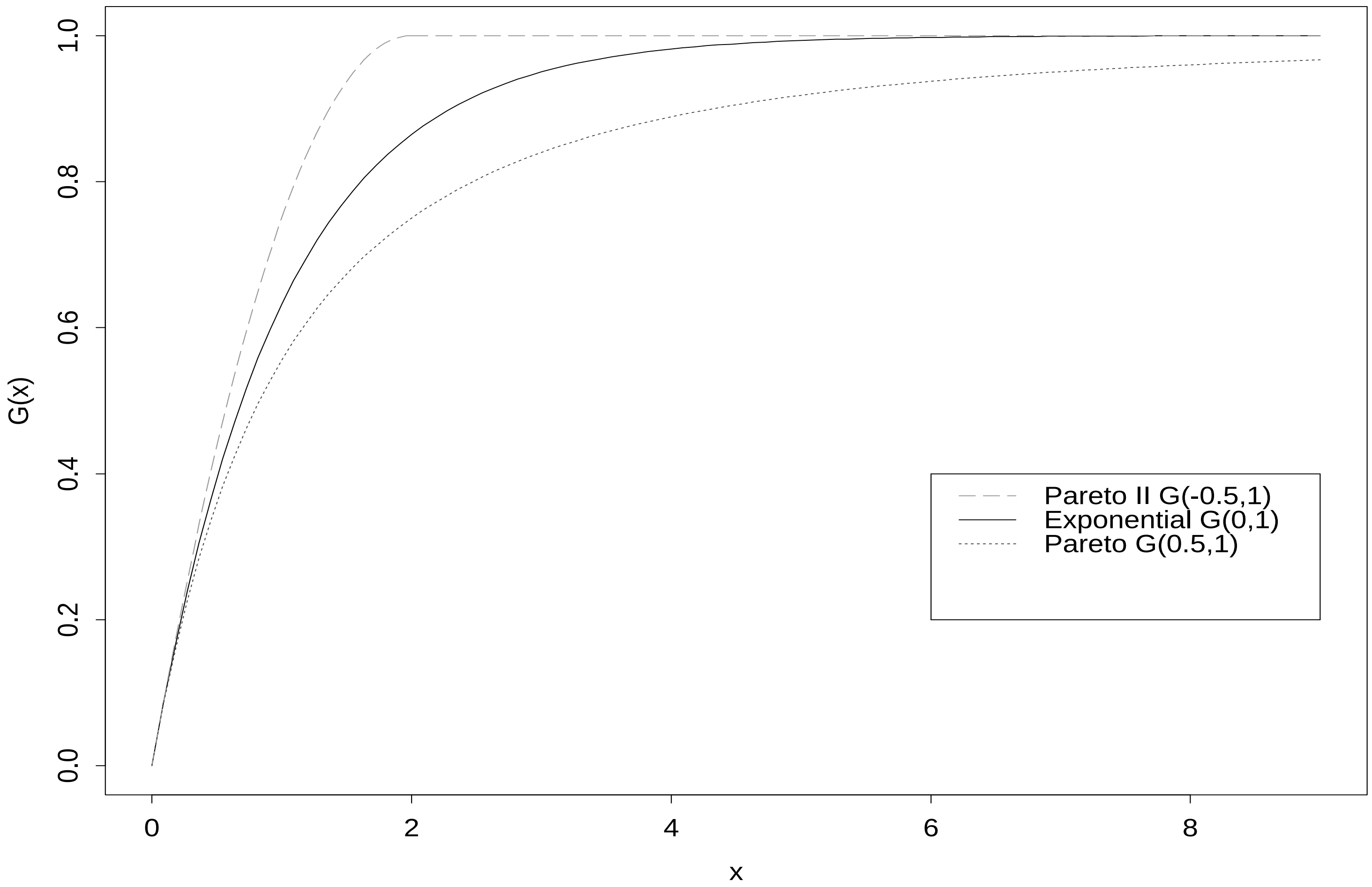
where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

This subsumes:

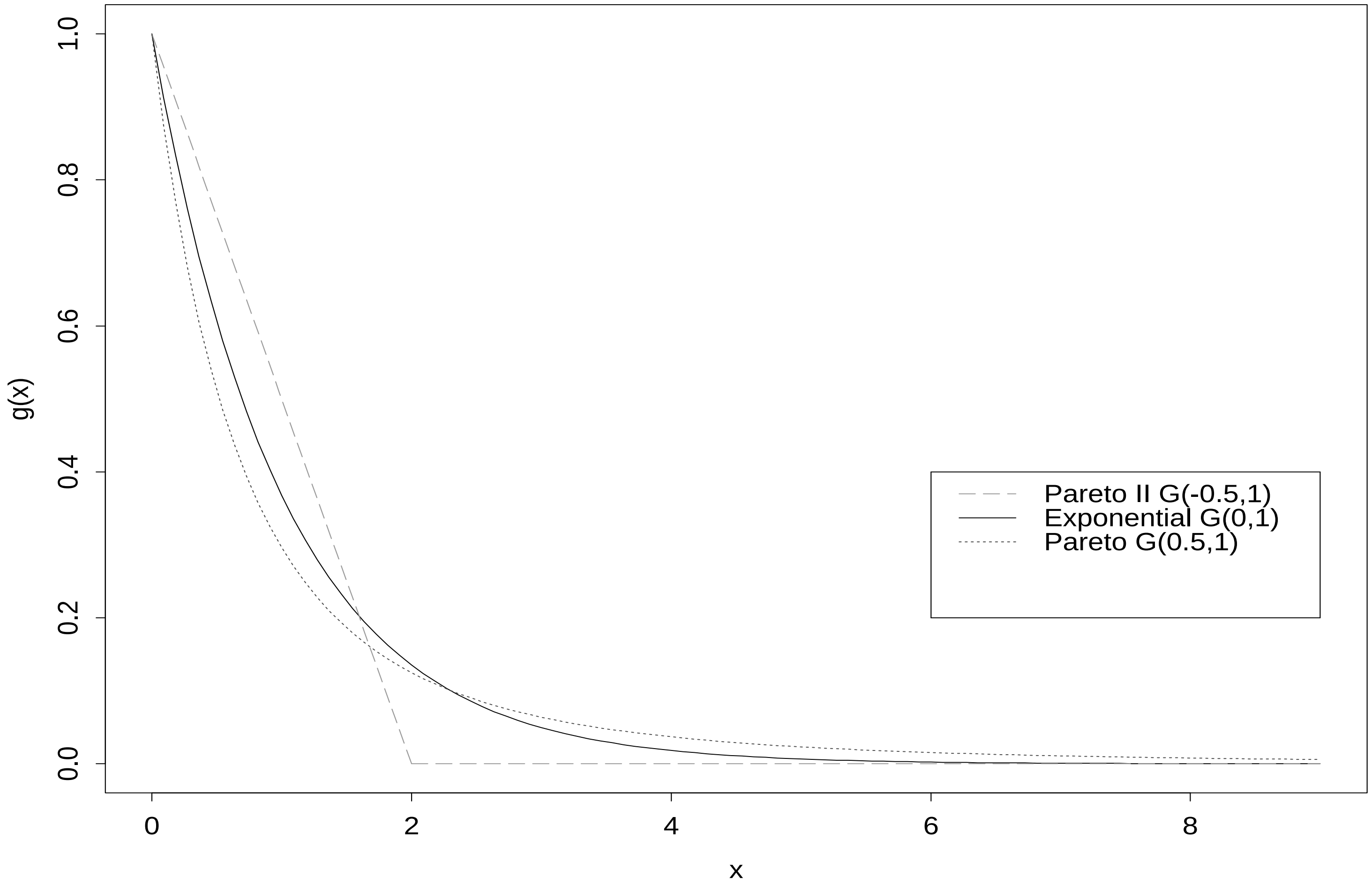
$\xi > 0$	Pareto (reparametrized version)
$\xi = 0$	exponential
$\xi < 0$	Pareto type II.

Moments. For $\xi > 0$ distribution is heavy tailed. $E\left(X^k\right)$ does not exist for $k \geq 1/\xi$.

GPD: distribution functions for various ξ



GPD: densities for various ξ



J2. POT Method: Theoretical Foundations

The excess distribution: Given that a loss exceeds a high threshold, by how much can the threshold be exceeded?

Let u be the high threshold and define the excess distribution above the threshold u to have the df

$$F_u(x) = P\left(X - u \leq x \mid X > u\right) = \frac{F(x+u) - F(u)}{1 - F(u)},$$

for $0 \leq x < x_F - u$ where $x_F \leq \infty$ is the right endpoint of F.

Extreme value theory suggests the GPD is a natural approximation for this distribution.

Examples

1. Exponential. $F(x) = 1 - e^{\lambda x}$, $\lambda > 0$, $x \geq 0$.

$$F_u(x) = F(x)\,, \quad x \geq 0\,.$$

The "lack–of–memory" property.

2. GPD. $F(x) = G_{\xi,\beta}(x)$.

$$F_u(x) = G_{\xi,\beta+\xi u}(x)\,,$$

where $0 \leq x < \infty$ if $\xi \geq 0$ and $0 \leq x < -\frac{\beta}{\xi} - u$ if $\xi < 0$.

The excess distribution of a GPD remains a GPD with the same shape parameter; only the scaling changes.

Asymptotics of Excess Distribution

Theorem. (Pickands–Balkema–de Haan (1974/75)) We can find a function $\beta(u)$ such that

$$\lim_{u \to x_F} \sup_{0 \le x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0 \,,$$

if and only if $F \in \text{MDA}\,(H_\xi)$, $\xi \in \mathbb{R}$.

Essentially all the common continuous distributions used in risk management or insurance mathematics are in MDA (H_ξ) for some value of ξ, as we will see below.

Exploiting Pickands–Balkema–de Haan

"For a wide class of distributions, the distribution of the excesses over high thresholds can be approximated by the GPD."

This result suggests we choose u high and assume the limit result is more or less exact

$$F_u(x) \approx G_{\xi,\beta}(x)\,,$$

for some ξ and β. To estimate these parameters we fit the GPD to the excess amounts over the threshold u. Standard properties of maximum likelihood estimators apply if $\xi > -0.5$.

To implement the POT method we must choose a suitable threshold u. There are data–analytic tools (e.g. mean excess plot) to help us here, although later simulations will suggest that inference is often robust to choice of threshold.

When does $F \in$ MDA (H_ξ) hold?

1. Fréchet Case: ($\xi > 0$)

Gnedenko (1943) showed that for $\xi > 0$

$$F \in \text{MDA}\,(H_\xi) \iff 1 - F(x) = x^{-1/\xi} L(x)\,,$$

for some slowly varying function $L(x)$.

A function L on $(0, \infty)$ is slowly varying if
$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1\,, \quad t > 0\,.$

Summary:

If the tail of the df F decays like a power function, then the distribution is in MDA (H_ξ) for $\xi > 0$.

When does $F \in$ MDA (H_ξ) hold? (II)

Examples of Fréchet case: Heavy-tailed distributions such as Pareto, Burr, loggamma, Cauchy and t–distributions as well as various mixture models. Not all moments are finite.

2. Gumbel Case: $F \in$ MDA (H_0)

The characterization of this class is more complicated. Essentially it contains distributions whose tails decay roughly exponentially and we call these distributions light–tailed. All moments exist for distributions in the Gumbel class.

Examples are the Normal, lognormal, exponential and gamma.

J3. Estimating Tails of Distributions

R Smith (1987) proposed a tail estimator based on GPD approximation to excess distribution. Let $N_u = \sum_{i=1}^{n} 1_{\{\{X_i > u\}\}}$ be the random number of exceedances of u from iid sample $X_1, \ldots, X_n$.

Note that for $x > u$ we may write $\overline{F}(x) = \overline{F}(u)\overline{F}_u(x-u)$.

We estimate $\overline{F}(u)$ empirically by N_u/n and and $\overline{F}_u(x-u)$ using a GPD approximation to obtain the tail estimator

$$\widehat{\overline{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}} ;$$

this estimator is only valid for $x > u$. A high u reduces bias in estimating excess function. A low u reduces variance in estimating excess function and $F(u)$.

Estimating Quantiles in Tail

Recall the qth quantile of F

$$x_q = F^{\leftarrow}(q) = \inf\{x \in \mathbb{R} : F(x) \geq q\}\,.$$

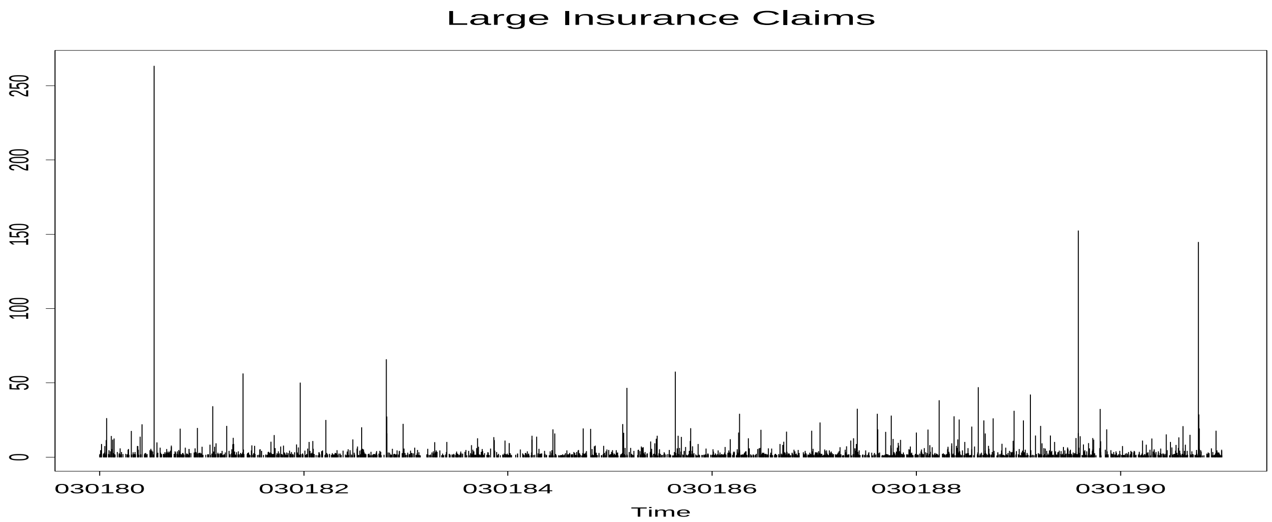
Suppose $x_q > u$ or equivalently $q > F(u)$. By inverting the tail estimation formula we get

$$\widehat{x}_q = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{n}{N_u}(1-q) \right)^{-\hat{\xi}} - 1 \right)\,.$$

Asymmetric confidence interval for x_q can be constructed using profile likelihood method.

J4. Danish Fire Loss Example

The Danish data consist of 2167 losses exceeding one million Danish Krone from the years 1980 to 1990. The loss figure is a total loss for the event concerned and includes damage to buildings, damage to contents of buildings as well as loss of profits. The data have been adjusted for inflation to reflect 1985 values.

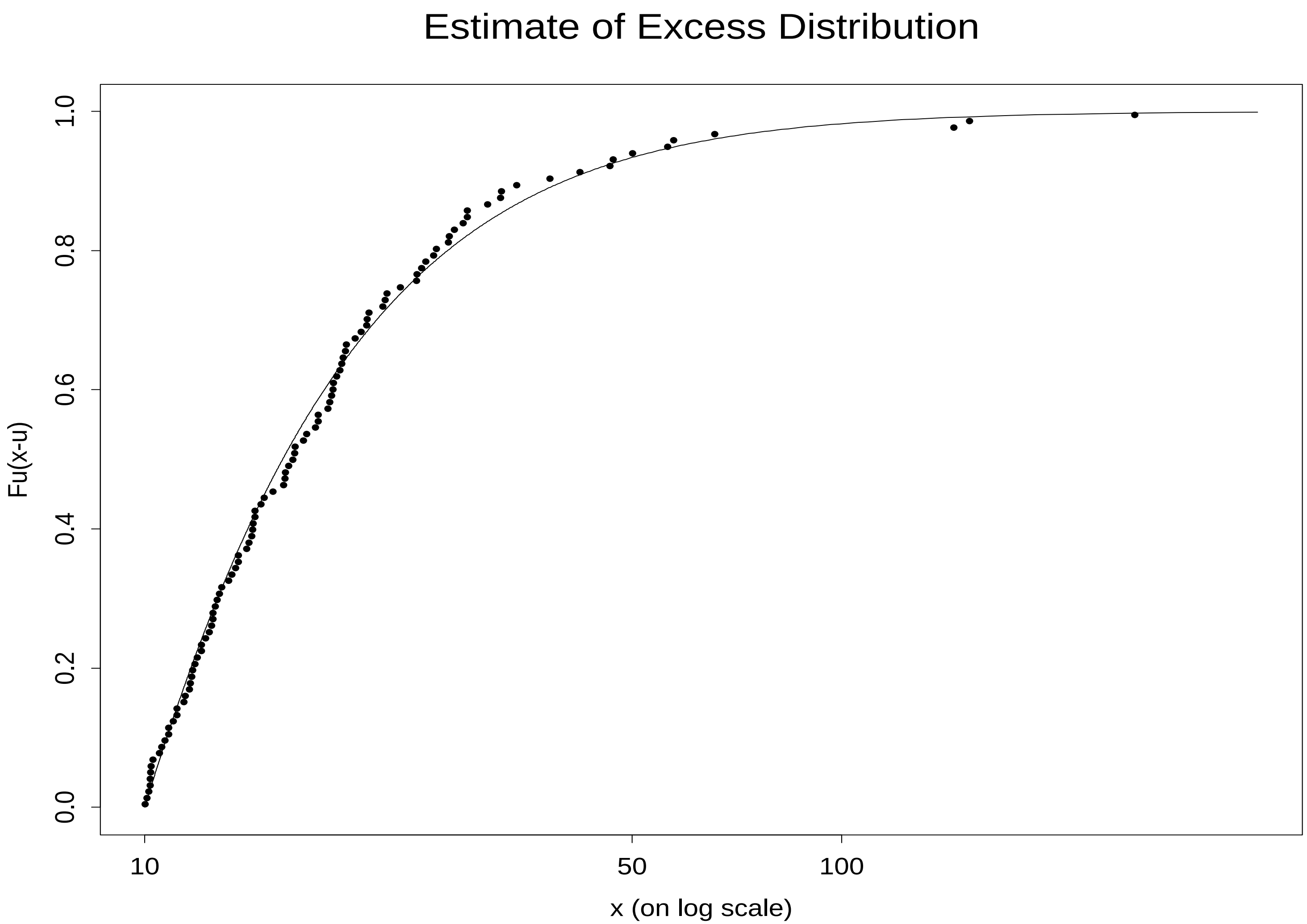


EVIS POT Analysis

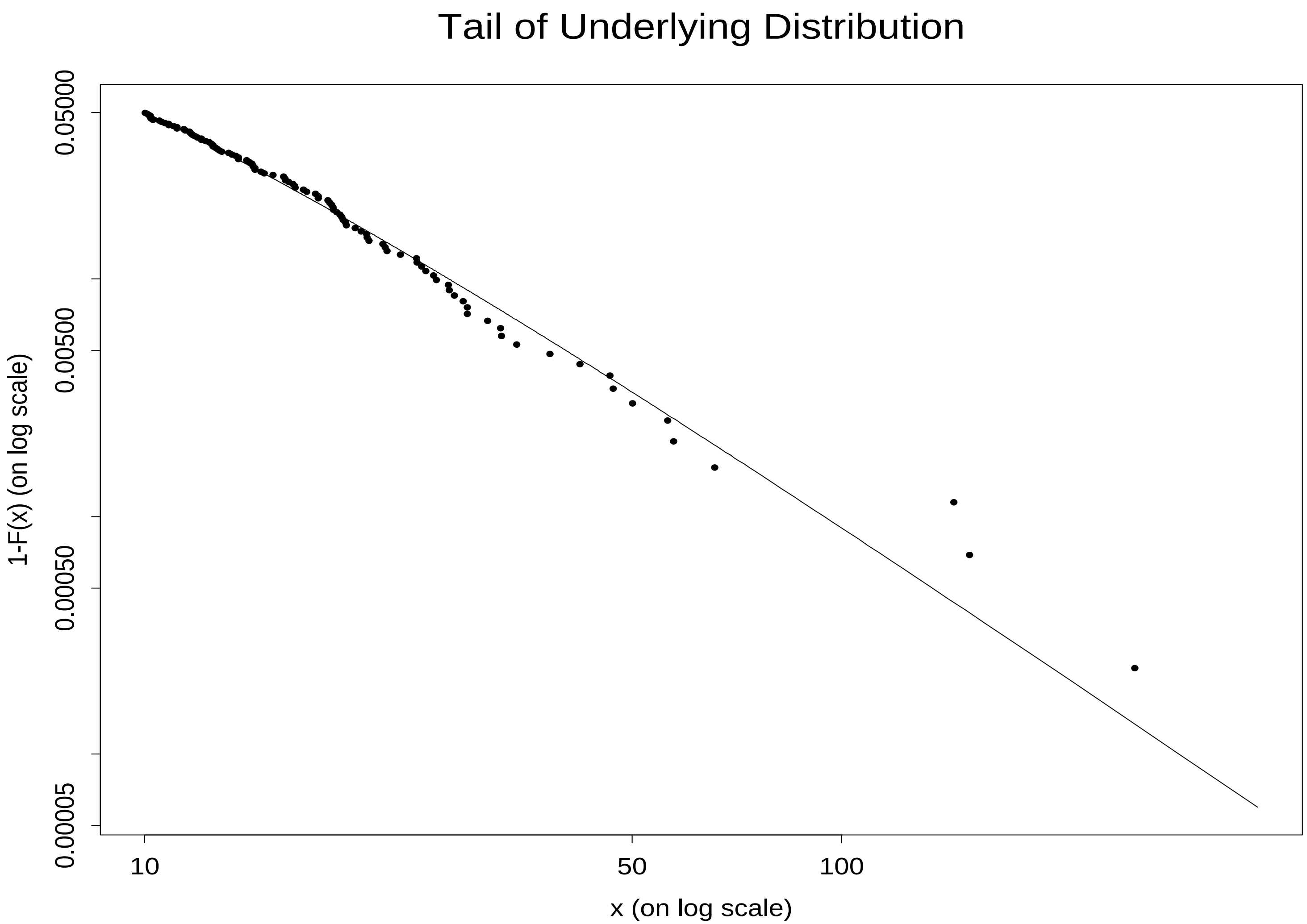
```
> out <- gpd(danish,10)
> out
$n:
[1] 2167

$data:
  [1]  11.37482  26.21464  14.12208
  [4]  11.71303  12.46559  17.56955
  [7]  13.62079  21.96193 263.25037
...etc...
[106] 144.65759  28.63036  19.26568
[109]  17.73927

$threshold:
[1] 10

$p.less.thresh:
[1] 0.9497

$n.exceed:
[1] 109
```

```
$par.ests:
        xi     beta
 0.4969857 6.975468

$par.ses:
        xi    beta
 0.1362838 1.11349

$varcov:
             [,1]        [,2]
[1,]  0.01857326 -0.08194611
[2,] -0.08194611  1.23986096

$information:
[1] "observed"

$converged:
[1] T

$nllh.final:
[1] 374.893

$
```

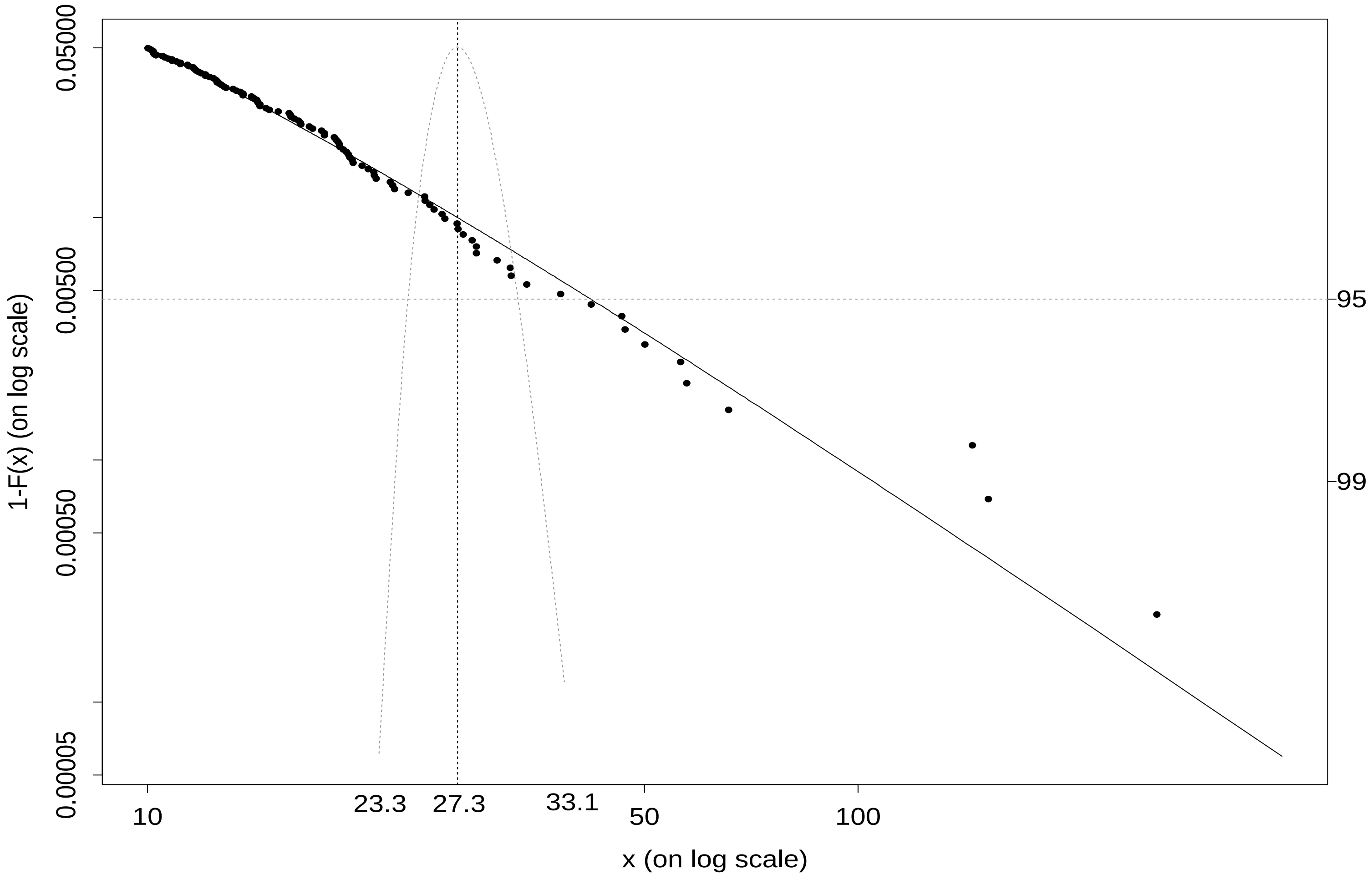
Estimating Excess df



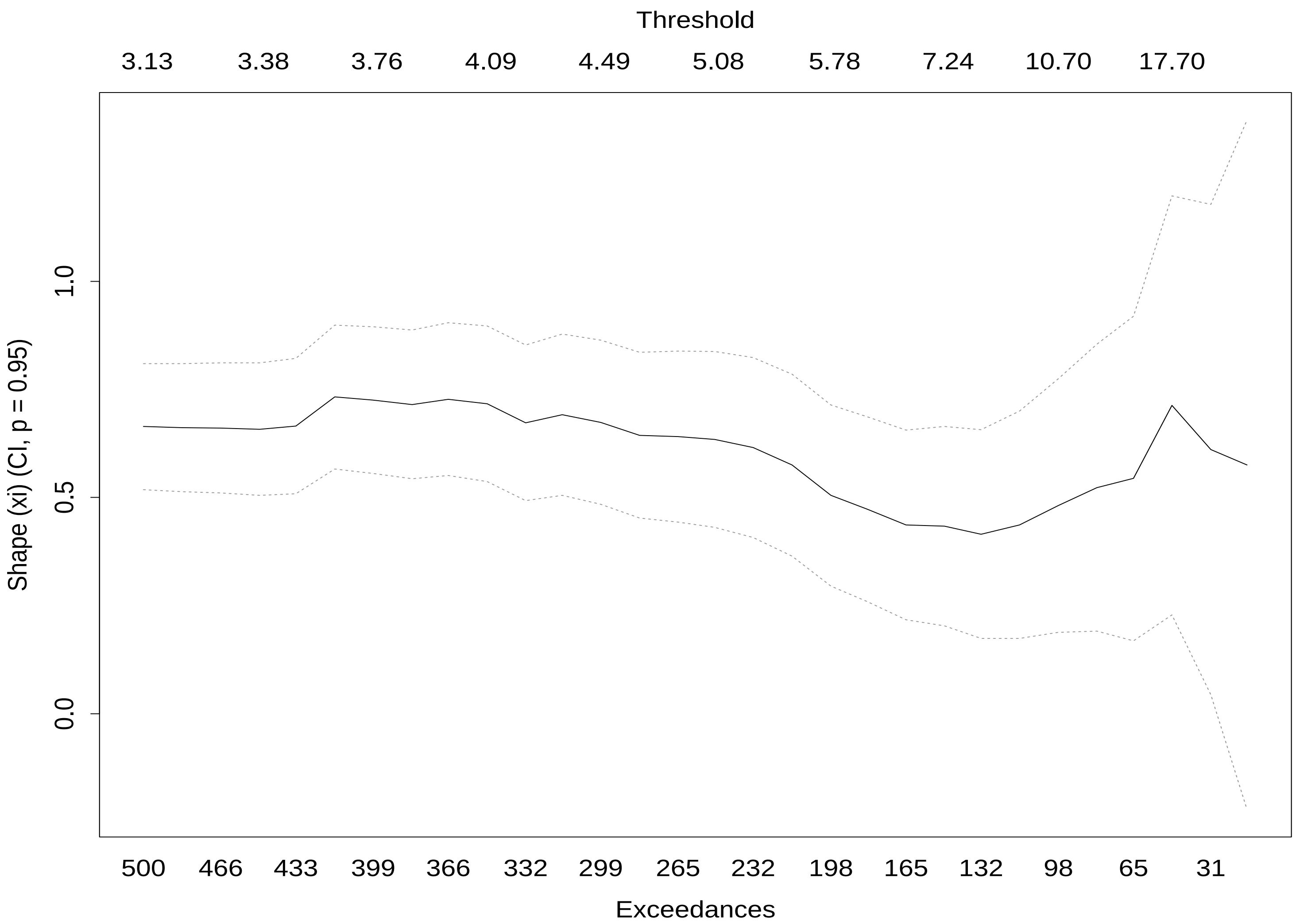
Estimating Tail of Underlying df



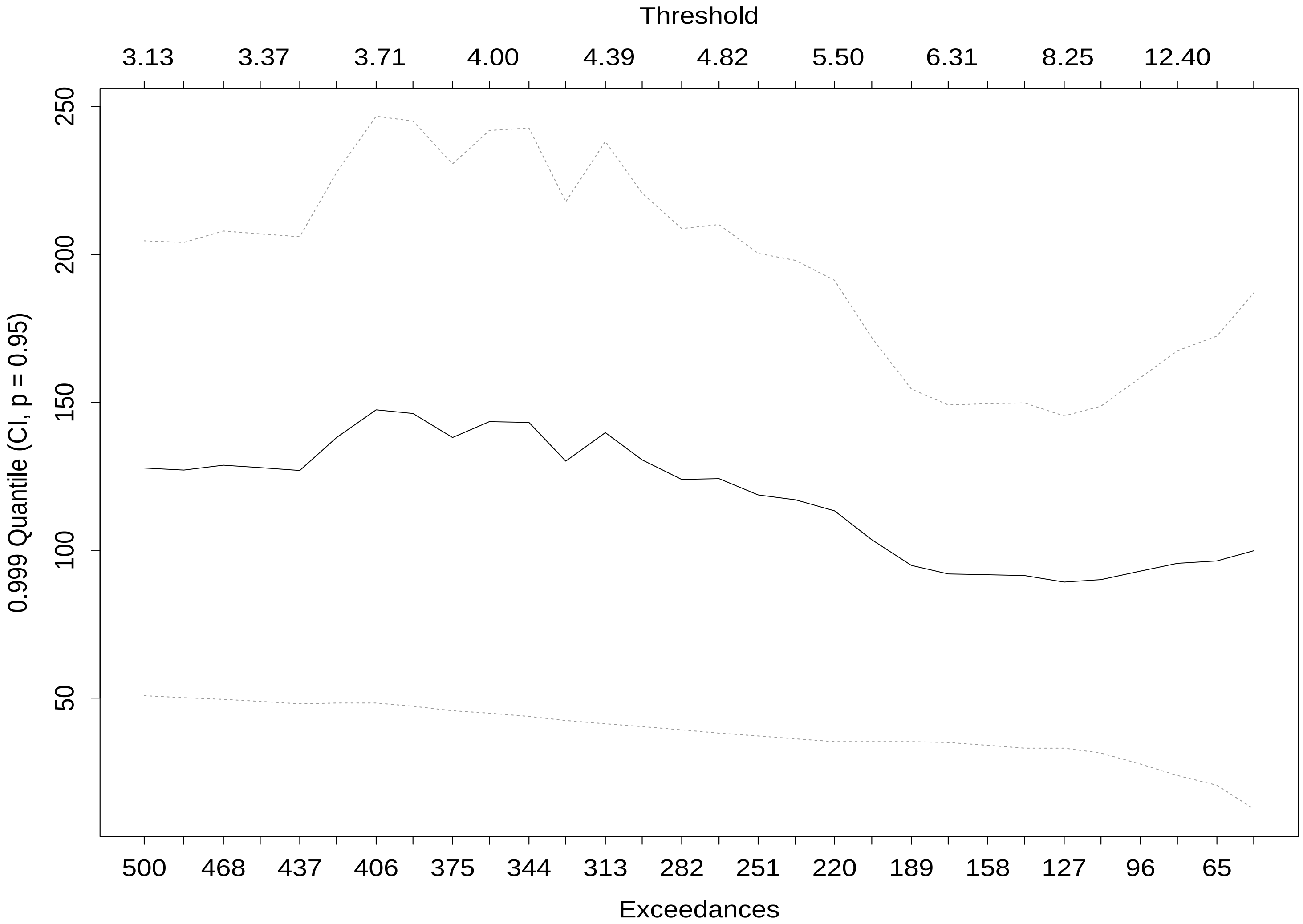
Estimating a Quantile (99%)



Varying the Threshold I



Varying the Threshold II



J5. Expected Shortfall and Mean Excess Plot

The mean excess function of a rv X is

$$e(u) = E(X - u \mid X > u).$$

It is the mean of the excess distribution function above the threshold u expressed as a function of u.

Our Model Assumption:

Excess losses over threshold u are exactly GPD with $\xi < 1$, i.e. $X - u \mid X > u \sim \text{GPD}(\xi, \beta)$. It is easily shown that for any higher threshold $v \geq u$

$$e(v) = E(X - v \mid X > v) = \frac{\beta + \xi(v - u)}{1 - \xi},$$

so that mean excess function is linear in v above u.

Sample Mean Excess Plot

The sample mean excess plot estimates $e(u)$ in the region where we have data:

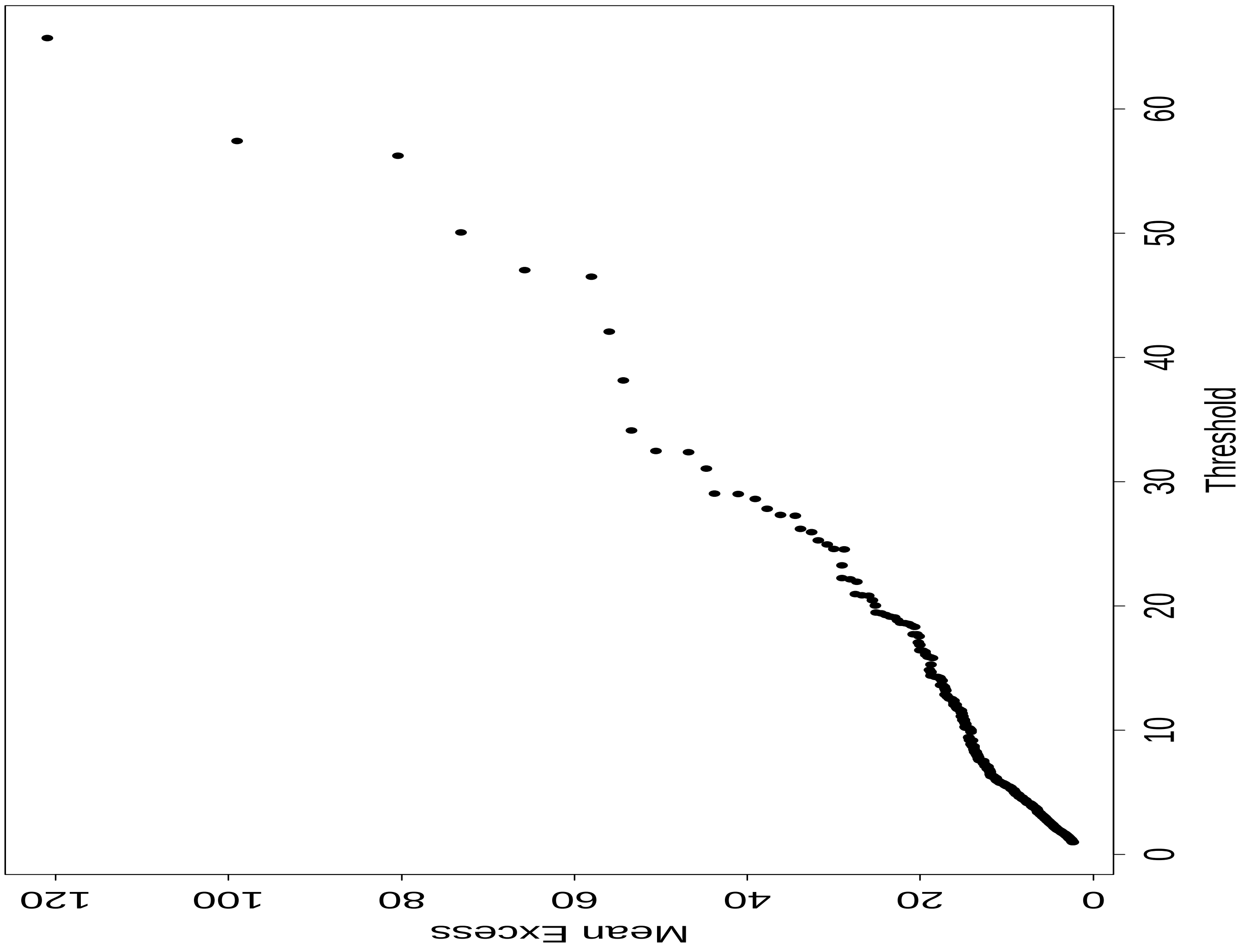
$$e_n(u) = \frac{\sum_{i=1}^{n}(X_i - u)^+}{\sum_{i=1}^{n} 1_{\{X_i > u\}}},$$

We seek a threshold u, above which the plot is roughly linear.

If we can find such a threshold, the result of Pickands-Balkema-De Haan could be applied above that threshold.

Note that the plot is erratic for large u, when the averaging is over very few excesses. It is often better to omit these from the plot.

Mean Excess Plot for Danish Data



Expected Shortfall: Estimation II

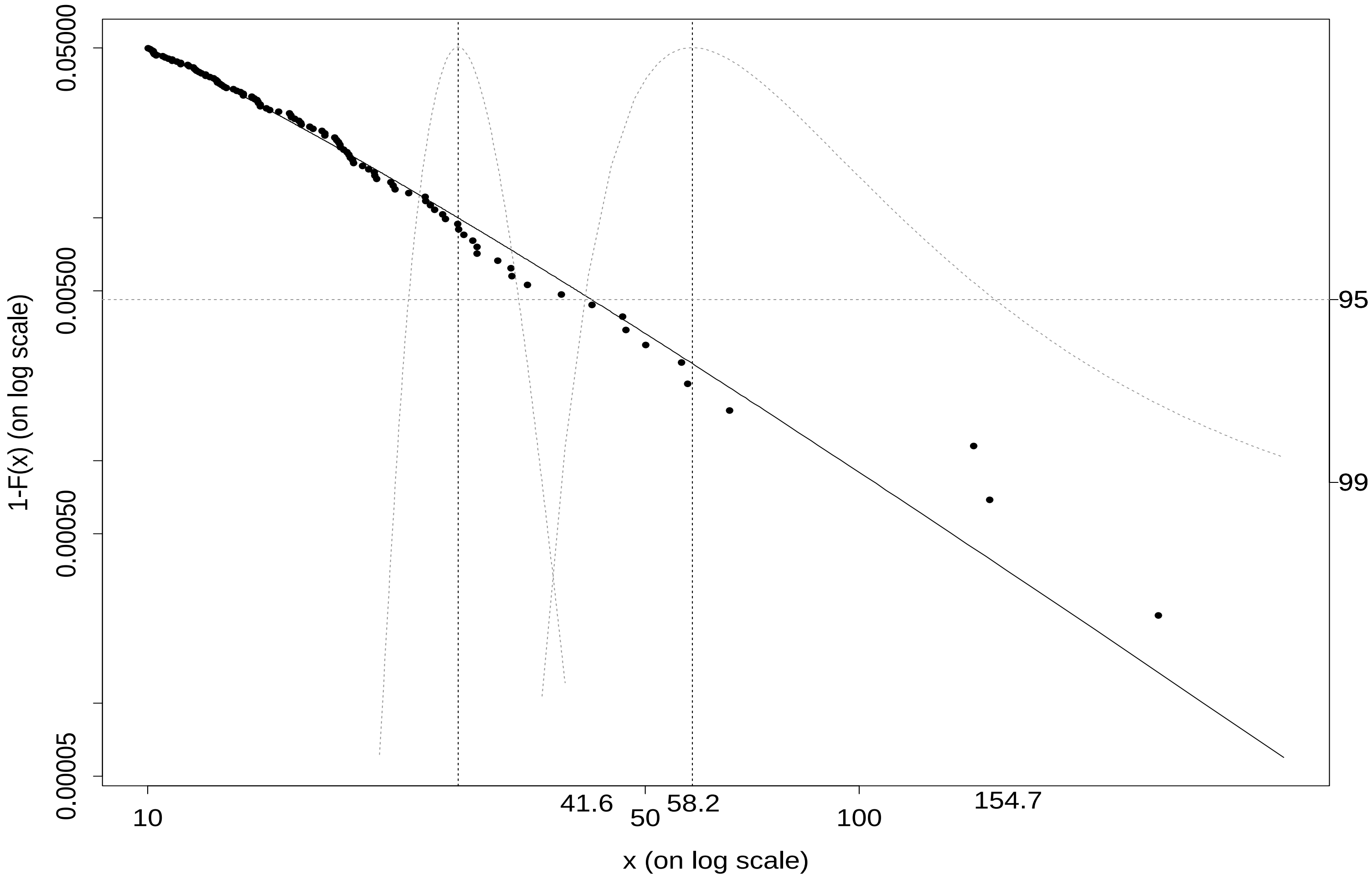
Now observe that for $x_q > u$

$$\begin{aligned} ES_q(X) &= E\left(X \mid X > x_q\right) \\ &= x_q + E\left(X - x_q \mid X > x_q\right) \\ &= x_q + \frac{\beta + \xi(x_q - u)}{1 - \xi}. \end{aligned}$$

This yields the estimator

$$\widehat{ES}_q(X) = \hat{x}_q \left(\frac{1}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{(1 - \hat{\xi})\hat{x}_q} \right).$$

Estimates of 99% VaR and ES (Danish Data)



K. Advanced Topics in EVT

1. Efficient Quantile Estimation with POT
2. The POT Method with Dependent Data
3. Dynamic EVT in Time Series Framework
4. An Example with S&P Data
5. VaR Estimation and Backtesting
6. Var for Longer Time Horizons – Scaling Rules

K1. Efficient Quantile Estimation with POT

Estimation of quantiles with POT is a more efficient method than simple empirical quantile estimation. The latter is often used in the historical simulation approach, but gives poor estimates when we are estimating at the edge of the sample.

Recall that we can compare the efficiency of two quantile estimators by comparing their mean squared errors (MSE). If $\widehat{x}_q$ is an estimator of x_q then

$$\begin{aligned} \text{MSE}\left(\widehat{x}_q\right) &= E\left(\left(\widehat{x}_q - x_q\right)^2\right) \\ &= \operatorname{var}\left(\widehat{x}_q\right) + E\left(\widehat{x}_q - x_q\right)^2 . \end{aligned}$$

Good estimators trade variance of against bias to give small MSE.

Comparison of Estimators

Take ordered data $X_{(1)} > \ldots > X_{(n)}$ (no ties) and place threshold u at an order statistic: $u = X_{(k+1)}$.

We emphasize dependence of POT estimator on choice of k by writing

$$\widehat{x}_{q,k} = X_{(k+1)} + \frac{\hat{\beta}_k}{\hat{\xi}_k}\left(\left(\frac{n}{k}(1-q)\right)^{-\hat{\xi}_k} - 1\right),$$

where $k \in \{j \in \mathbb{N} : j \geq n(1-q)\}$.

The empirical quantile estimator is $\widehat{x}_q^E = X_{([n(1-q)]+1)}$.

Example. $n = 1000$ implies $\widehat{x}_{0.995}^E = X_{(6)}$.

We compare $\mathrm{MSE}\left(\widehat{x}_{q,k}\right)$ with $\mathrm{MSE}\left(\widehat{x}_q^E\right)$.

Simulation Study

For various underlying F, various sample sizes n and various quantile probabilities q we compare the MSEs of these estimators. MSEs are estimated by Monte Carlo, i.e. repeated simulation of random samples from F.

Examples

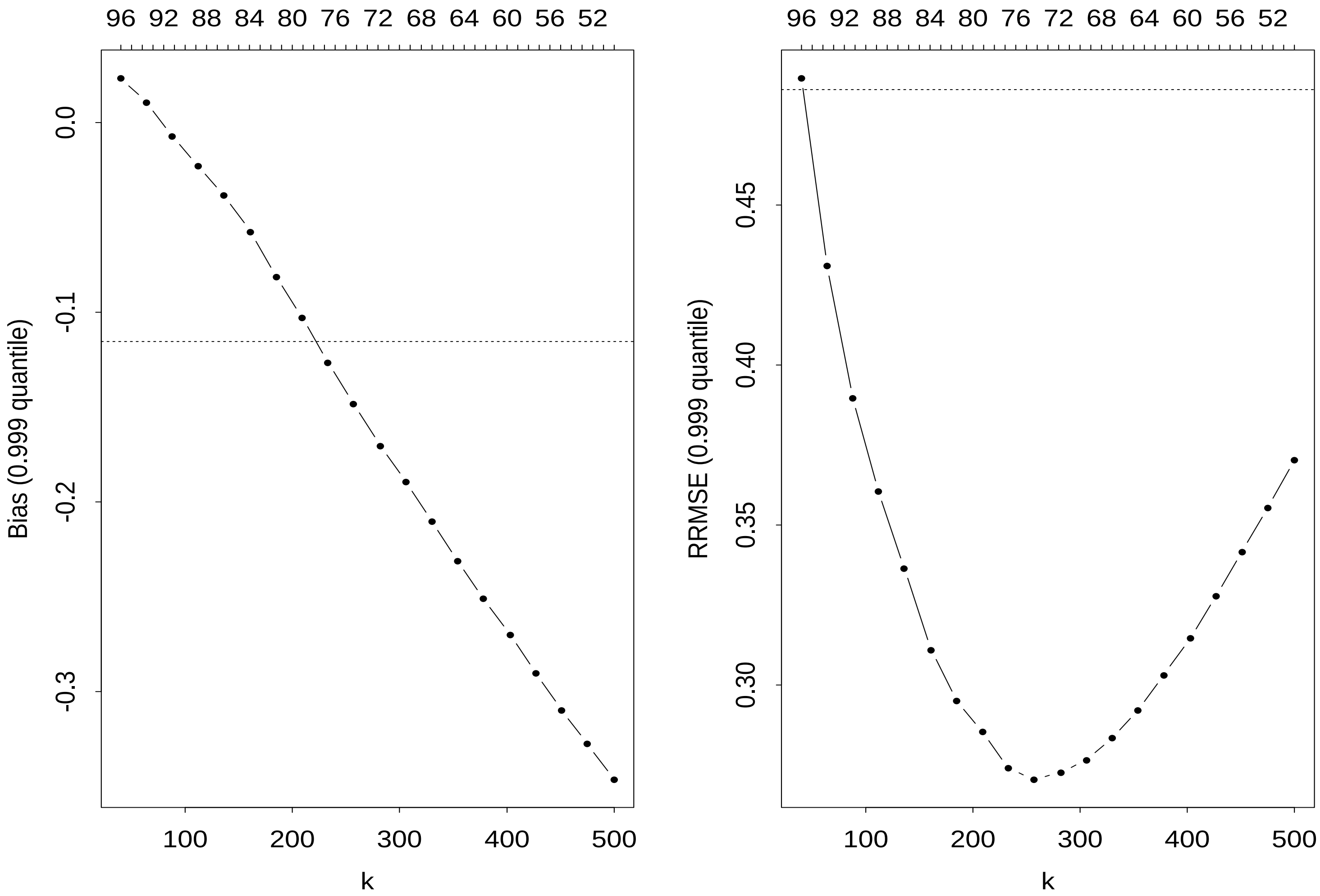
Hard: t–distribution, $n = 1000$, $q = 0.999$.
Easy: normal distribution $n = 1000$, $q = 0.95$.

We will actually compare

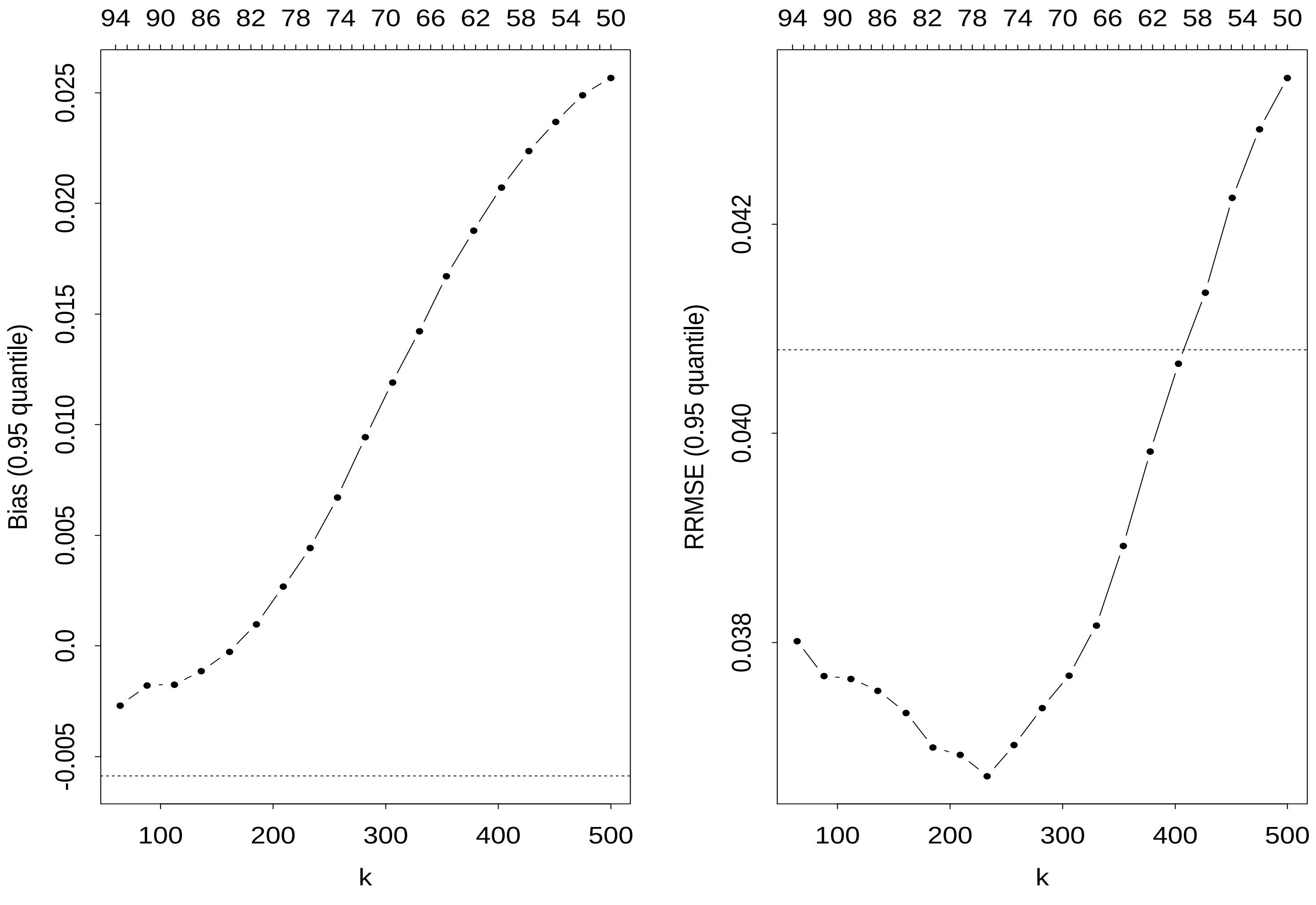
$$\text{RRMSE}\left(\hat{x}_q\right) = \frac{\sqrt{\text{MSE}\left(\hat{x}_q\right)}}{x_q}$$

to express error relative to original units.

0.999 Quantile of t with $\nu = 2$



0.95 Quantile of Standard Normal



K2. Statistical Implications of Dependence

If we believe we have a (strictly) stationary time series with a stationary distribution F in the MDA of an extreme value distribution, then we can still apply the POT method and attempt to approximate the excess distribution $F_u(x)$ by a GPD for some high threshold u.

Although the marginal distribution of excesses may be approximately GPD, the joint distribution is unknown. We form the likelihood by making the simplifying assumption of independent excesses.

We can expect our estimation procedure to deliver consistent parameter estimates, but standard errors and confidence intervals may be over-optimistically small. Dependent samples carry less information about extreme events than independent samples of the same size.

Other Possibilities

- Use statistical estimation method for GPD parameters which does not implicitly assume independence of the excesses, such as probability weighted moments. However this method does not deliver standard errors.

- Attempt to make the excesses more independent by the technique of declustering and then use ML estimation. We identify clusters of exceedances and reduce each cluster to a single representative such as the cluster maximum.

K3. EVT in a Time Series Framework

We assume (negative) returns follow stationary time series of the form

$$X_t = \mu_t + \sigma_t Z_t \,.$$

Dynamics of conditional mean μ_t and conditional volatility σ_t are given by an AR(1)-GARCH(1,1) model:

$$\begin{aligned}
\mu_t &= \phi X_{t-1} \,, \\
\sigma_t^2 &= \alpha_0 + \alpha_1 \left(X_{t-1} - \mu_{t-1}\right)^2 + \beta \sigma_{t-1}^2 \,,
\end{aligned}$$

with α_0, α_1, $\beta > 0$, $\alpha_1 + \beta < 1$ and $|\phi| < 1$.

We assume (Z_t) is strict white noise with $E(Z_t) = 0$ and $\mathrm{var}(Z_t) = 1$, but leave exact innovation distribution unspecified. Other GARCH-type models could be used if desired.

Dynamic EVT

Given a data sample $x_{t-n+1}, \ldots, x_t$ from (X_t) we adopt a two-stage estimation procedure. (Typically we take $n = 1000$.)

- We forecast μ_{t+1} and σ_{t+1} by fitting an AR–GARCH model with unspecified innovation distribution by pseudo-maximum-likelihood (PML) and calculating 1–step predictions.
  (PML yields consistent estimator of GARCH–parameters)

- We consider the model residuals to be iid realisations from the innovation distribution and estimate the tails of this distribution using EVT (GPD-fitting). In particular estimates of quantiles z_q and expected shortfalls $E\left[Z \mid Z > z_q\right]$ for the distribution of (Z_t) can be determined.

Risk Measures

Recall that we must distinguish between risk measures based on tails of conditional and unconditional distributions of the loss - in this case the negative return.

We are interested in the former and thus calculate risk measures based on the conditional distribuion $F_{[X_{t+1}|\mathcal{F}_t]}$.

For a one-step time horizon risk measure estimates are easily computed from estimates of z_q and $E\left[Z \mid Z > z_q\right]$ and predictions of μ_{t+1} and σ_{t+1} using

$$\begin{aligned}\mathrm{VaR}_q(X_{t+1}) &= \mu_{t+1} + \sigma_{t+1} z_q,\\ ES_q(X_{t+1}) &= \mu_{t+1} + \sigma_{t+1} E\left[Z \mid Z > z_q\right].\end{aligned}$$

Dynamic EVT II

Advantages of this approach

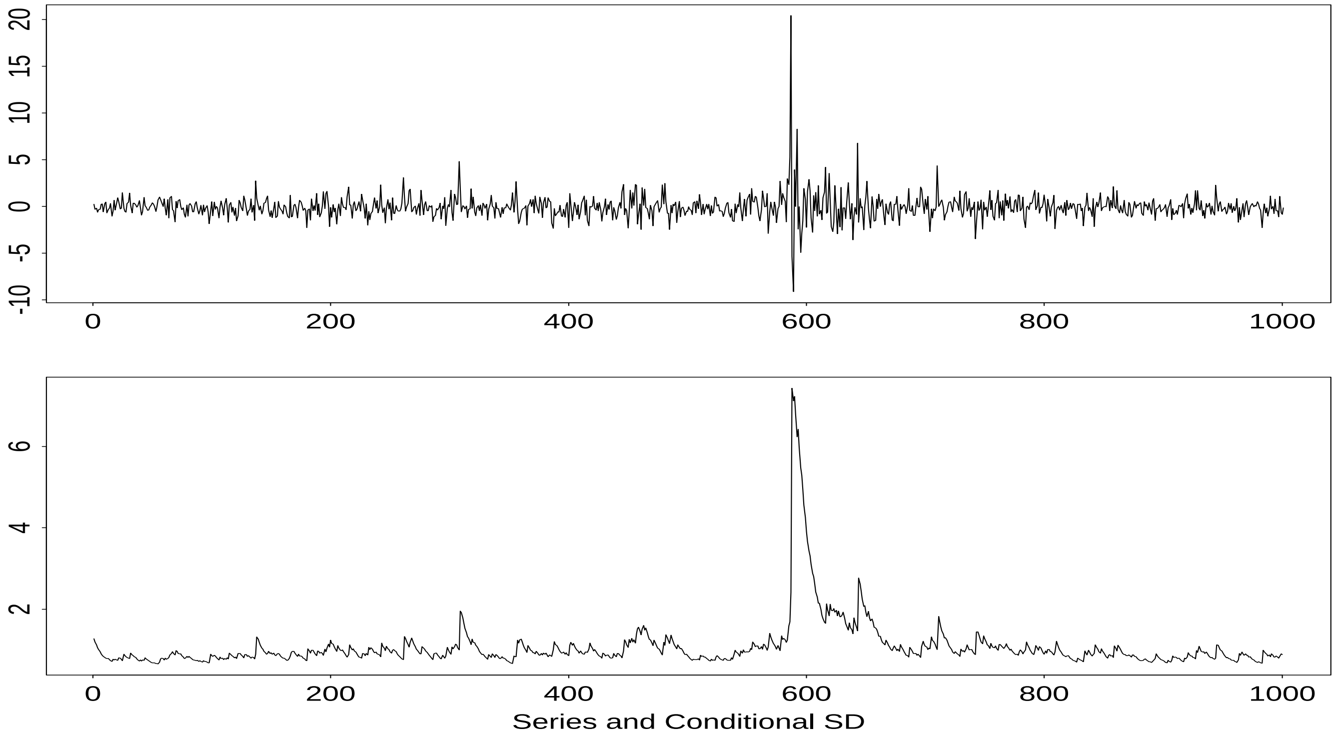
We model tails of innovation distribution explicitly, using methods which are supported by statistical theory. Residuals are approximately iid, so use of standard POT procedure is unproblematic.

Alternative Estimation Approaches.

(a) Assume (X_t) is GARCH process with normal innovations and fit by standard ML. In practice high quantiles are often underestimated.

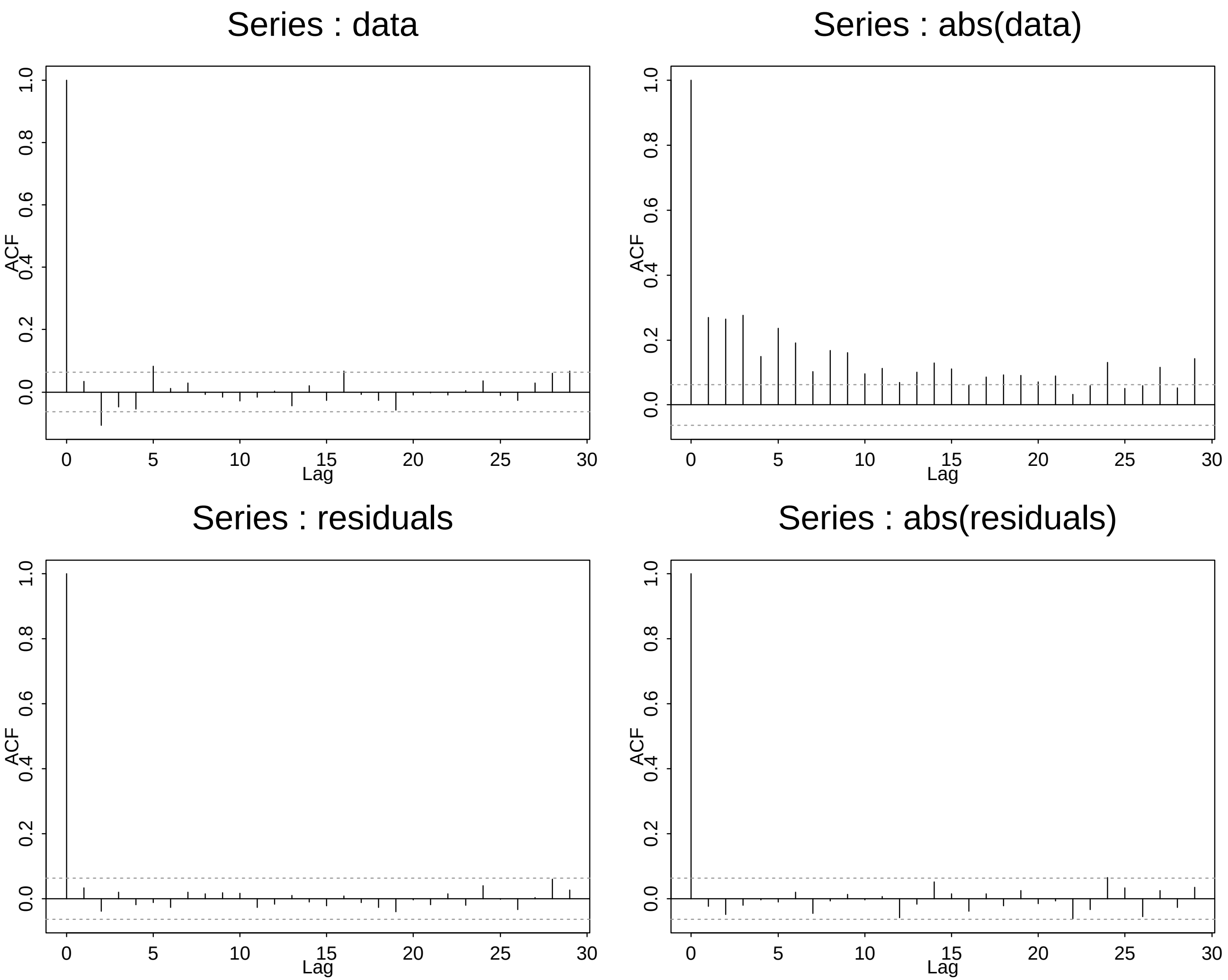
(b) Assume (X_t) is GARCH process with scaled t_ν–innovations. Use ML to estimate ν and GARCH–parameters at the same time.
In practice: this works much better but has some problems with asymmetric return series.

K4. Example with S&P Data

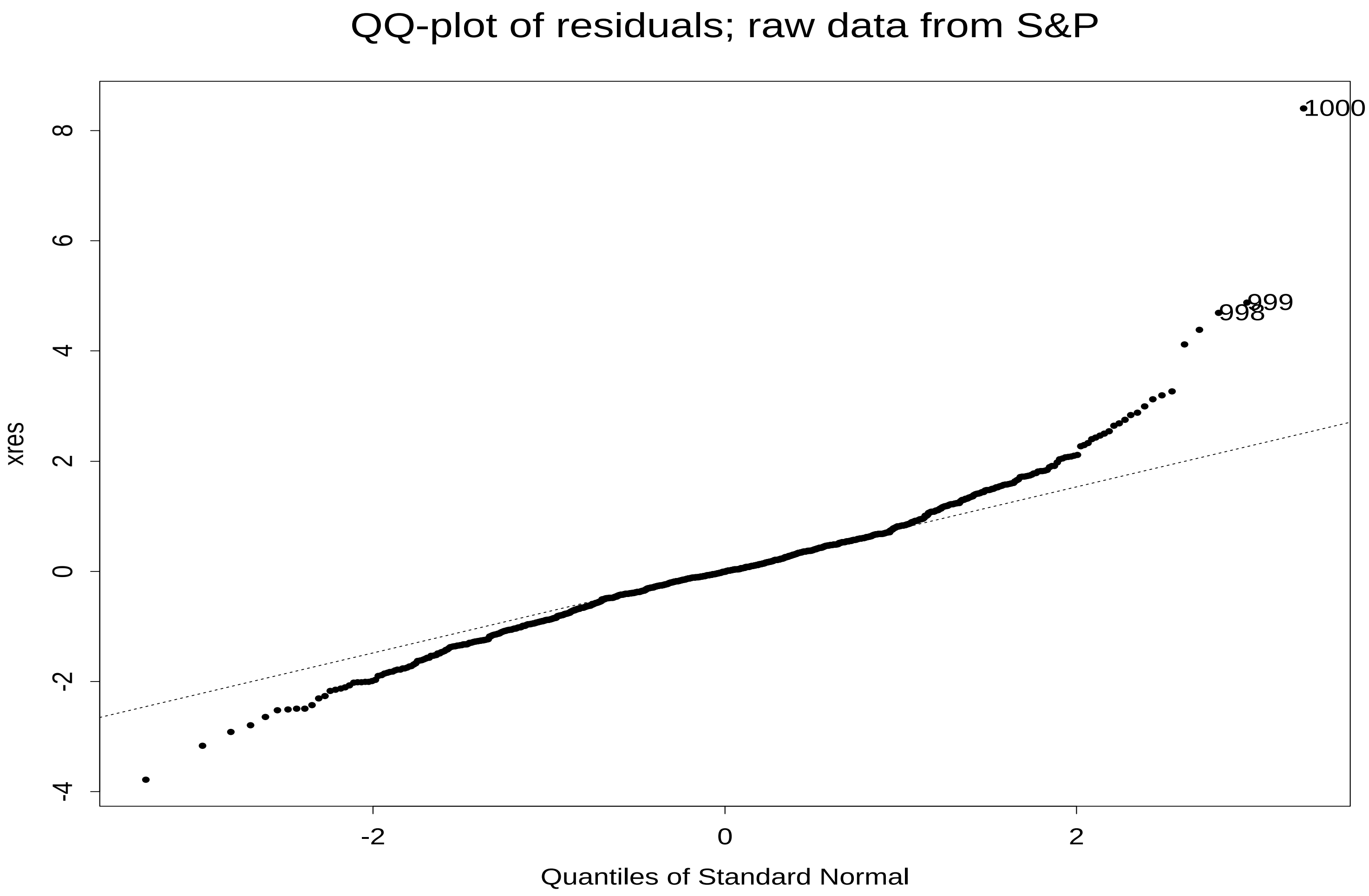


1000 day excerpt from series of negative log returns on Standard & Poors index containing crash of 1987.

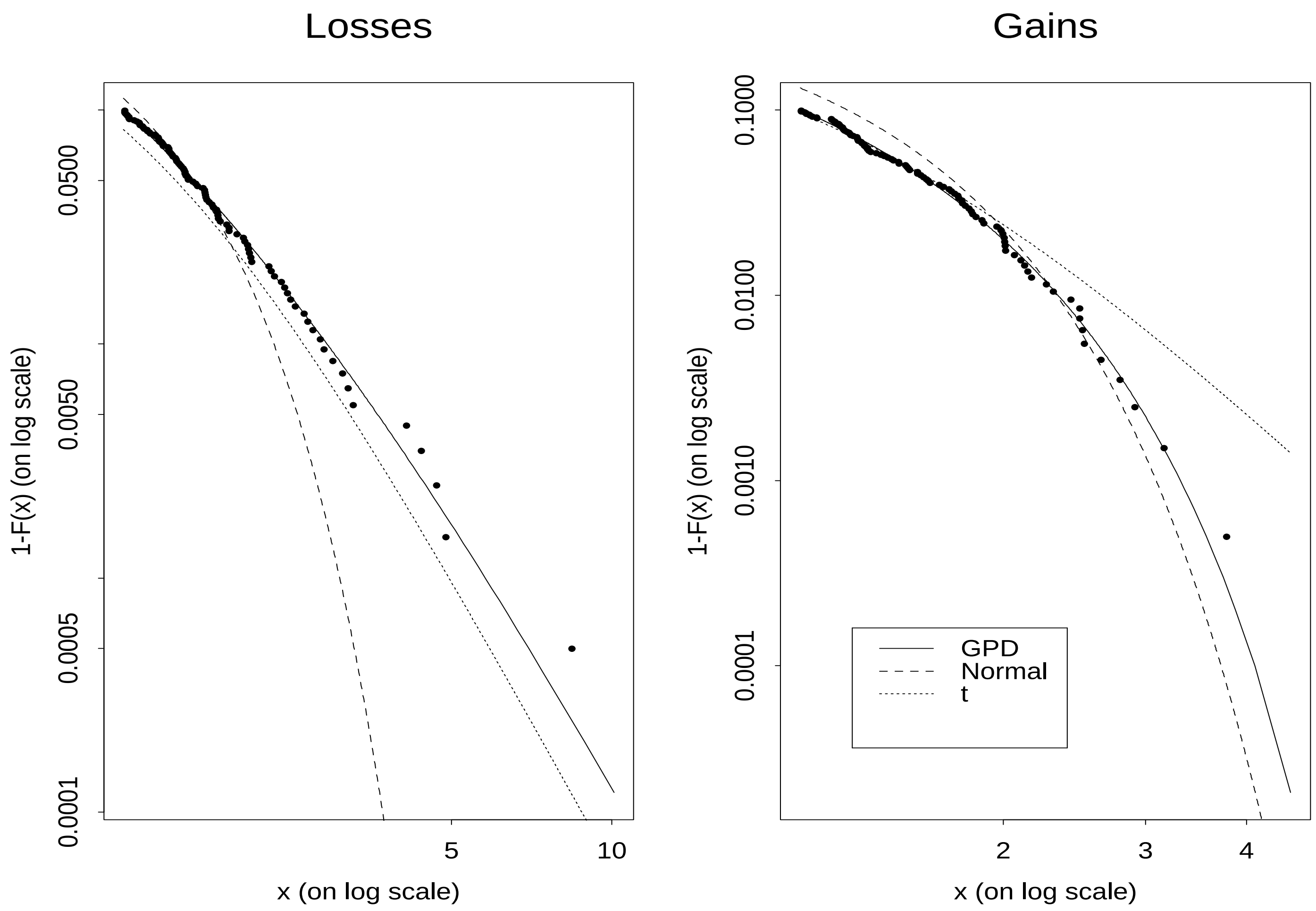
“Prewhitening” with GARCH



Heavy–Tailedness Remains



Comparison with Standard Conditional Distributions



K5. Backtesting

Eidgenössische Technische Hochschule Zürich — Department of Mathematics D-MATH

The ETH *Riskometer*

Market Risk Summary for Major Indices on 18/04/00

Dynamic Risk Measures

Index	VaR (95%)	ESfall (95%)	VaR (99%)	ESfall (99%)	Volatility
S&P 500	3.98	5.99	7.16	9.46	40.1
Dow Jones	3.66	5.43	6.47	8.47	37.4
DAX	3.08	4.21	4.89	6.12	29.3

- **VaR and ESfall** prognoses are estimates of potential daily losses expressed as percentages.
- **Volatility** is an annualized estimate expressed as a percentage; click on column heading for recent history.
- **Data** are kindly provided by Olsen & Associates.
- **Developers are** Alexander McNeil and Rüdiger Frey in the group for financial and insurance mathematics in the mathematics department of ETH Zürich.
- **Our methods**, which combine econometric modelling and extreme value theory, are described in our research paper; there are postscript and pdf versions.

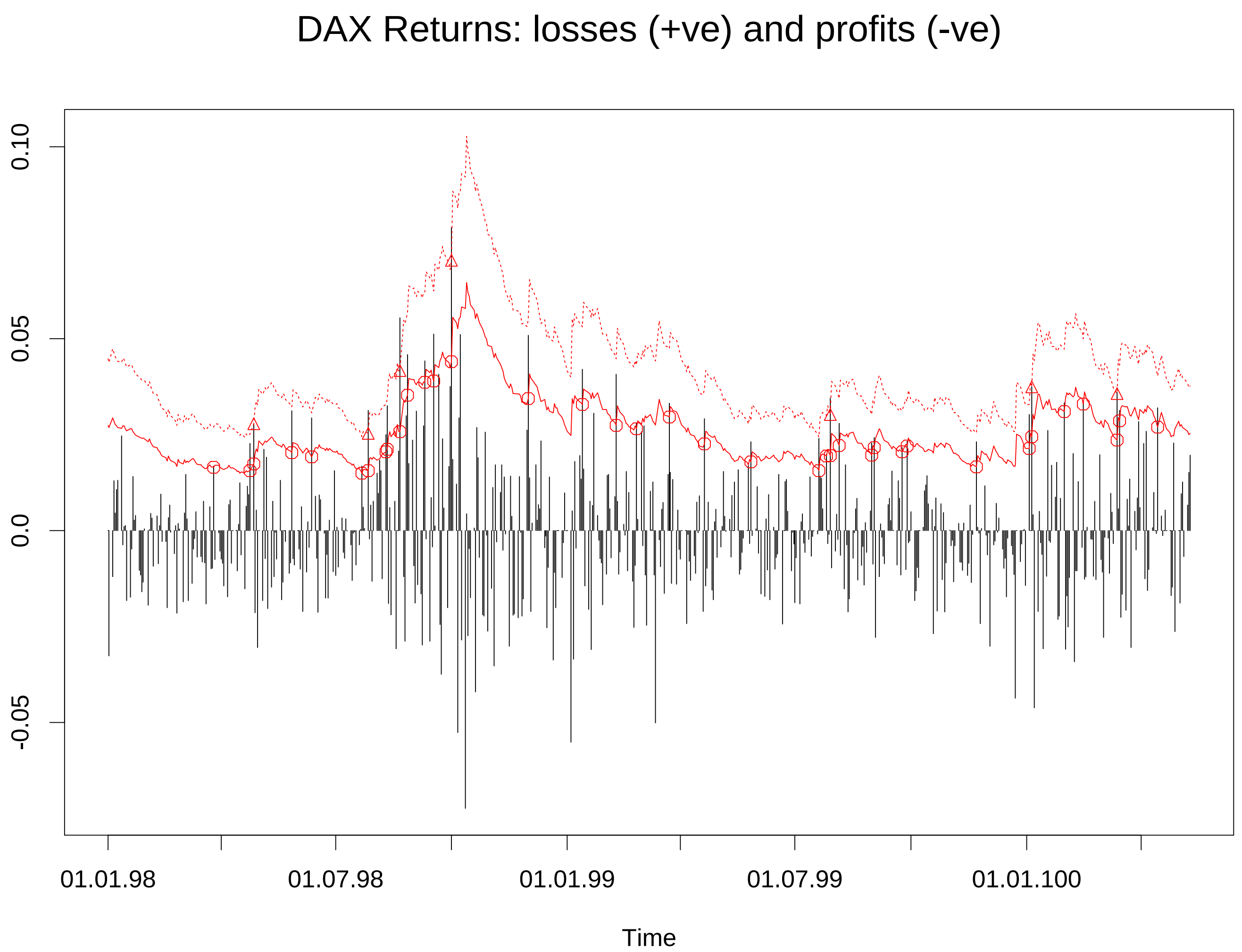
VaR Backtests & Violation Summary

- DAX backtest table or picture
- Dow Jones backtest table or picture
- S&P backtest table or picture

In all backtest pictures the 95% VaR is marked by a solid red line and the 99% VaR by a dotted red line. Circles and triangles indicate violation respectively.

Alexander McNeil (mcneil@math.ethz.ch)

Dynamic EVT: 95% and 99% VaR Predictions



Backtesting II – numbers of violations

	S&P	DAX
Length of Test	7414	5146
0.95 Quantile		
Expected	371	257
Conditional EVT	366 (0.41)	258 (0.49)
Conditional Normal	384 (0.25)	238 (0.11)
Conditional t	404 (0.04)	253 (0.41)
Unconditional EVT	402 (0.05)	266 (0.30)
0.99 Quantile		
Expected	74	51
Conditional EVT	73 (0.48)	55 (0.33)
Conditional Normal	104 (0.00)	74 (0.00)
Conditional t	78 (0.34)	61 (0.11)
Unconditional EVT	86 (0.10)	59 (0.16)

Remark: Performance of ES estimates even more sensitive to suitability of model in the tail region.

K6. Multi-day returns: Simulation of P&L

We adopt a Monte Carlo procedure and simulate from our dynamic model. We simulate iid noise from composite distribution made up of empirical middle and GPD tails.

$$\hat{F}_Z(z) = \begin{cases} \frac{k}{n}\left(1 + \xi_k^{(2)} \frac{|z - z_{(n-k)}|}{\beta_k^{(2)}}\right)^{-1/\xi_k^{(2)}} & \text{if } z < z_{(n-k)}, \\ \frac{1}{n}\sum_{i=1}^{n} 1_{\{z_i \le z\}} & \text{if } z_{(n-k)} \le z \le z_{(k+1)}, \\ 1 - \frac{k}{n}\left(1 + \xi_k^{(1)} \frac{z - z_{(k+1)}}{\beta_k^{(1)}}\right)^{-1/\xi_k^{(1)}} & \text{if } z > z_{(k+1)}. \end{cases}$$

For an h-day calculation we simulate 1000 (say) conditionally independent future paths $x_{t+1}, \ldots, x_{t+h}$ and compute simulated iid observations $x_{t+1} + \ldots + x_{t+h}$. Risk measures are estimated from simulated data.

Empirical Multi–day Results

Goal: assess performance and compare with "square root of time rule" (valid for iid normally distributed returns).

	S&P	DAX	BMW
$h = 10$; length of test	7405	5136	5136
0.95 Quantile			
Expected	370	257	257
Conditional EVT (h-day)	403	249	231
Square–root–of–time	623	318	315
0.99 Quantile			
Expected	74	51	51
Conditional EVT (h-day)	85	48	53
Square-root-of-time	206	83	70

Square root of time scaling does not seem sophisticated enough!
Note that formal statistical testing difficult because of overlapping returns.

How to model Operational Risk?

Paul Embrechts

Director RiskLab, Department of Mathematics, ETH Zurich

Member of the ETH Risk Center

Senior SFI Professor

http://www.math.ethz.ch/~embrechts

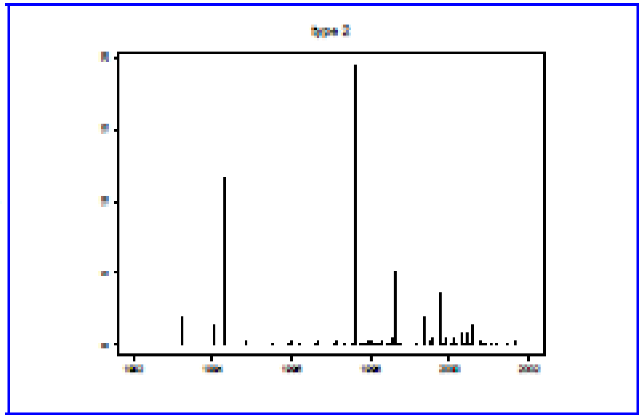
Risk Components (Basel II) now Basel III even Basel 3.5 ...

- **Credit Risk**
- **Market Risk**
- **Operational Risk**
- **Business Risk** ...

Operational Risk: The risk of loss resulting from inadequate or failed internal processes, people and systems or from external events. Including legal risk, but excluding strategic and reputational risk.

Loss Distribution Approach (LDA) within AMA-Framework

	RT_1 $\cdots$	RT_k	$\cdots$ RT_7	
BL_1 $\vdots$				
BL_i		$L_{i,k}^{T+1}$		
$\vdots$ BL_8				
				L^{T+1}



Internal, external, expert opinion data

Matrix structured loss data

A complicated stochastic structure

"Insurance Analytics"

$$
\left\{
\begin{array}{rcl}
L^{T+1} & = & \displaystyle\sum_{i=1}^{8} \sum_{k=1}^{7} L_{i,k}^{T+1} \\[2em]
L_{i,k}^{T+1} & = & \displaystyle\sum_{\ell=1}^{N_{i,k}^{T+1}} X_{i,k}^{\ell} \\[2em]
X_{i,k}^{\ell} & : & \text{loss severities} \\[1em]
N_{i,k}^{T+1} & : & \text{loss frequencies}
\end{array}
\right.
$$

together with left-censoring, reporting delays (IBNR-like), insurance cover, ...

The two relevant (regulatory) risk measures:

Value-at-Risk (VaR) and Expected Shortfall (ES)

$\mathrm{VaR}_p(X)$

For $p \in (0,1)$,

$$\mathrm{VaR}_p(X) = F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \geq p\}$$

$\mathrm{ES}_p(X)$

For $p \in (0,1)$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q \underset{(F\text{ cont.})}{=} \mathbb{E}\left[X | X > \mathrm{VaR}_p(X)\right]$$

Basel II - Guidelines

- **Risk measure:** VaR (= Value-at-Risk, a quantile)
- **Time horizon:** 1 year
- **Level:** 99.9% (1 in 1000 year event!)

(Extreme event!)

"Darwinism"

▶ **Otherwise:** Full methodological freedom (within LDA)

How to model Operational Risk …

… if you must!

- Discussion between “Yes we can” and “No you can’t”
- Banking versus Insurance:

  An example: Lausanne 2006 → BPV, EBK, FINMA …

- The record loss as of today: BoA’s 16.65 billion USD settlement with the DOJ (August 2014), of which14.54 billion USD corresponds to BCBS Event type “Suitability, disclosure and fiduciary” and Business Line “Trading and sales”
- One thing is for sure:

  Operational Risk is of paramount importance!

  But how reliably can it be quantitatively risk managed?

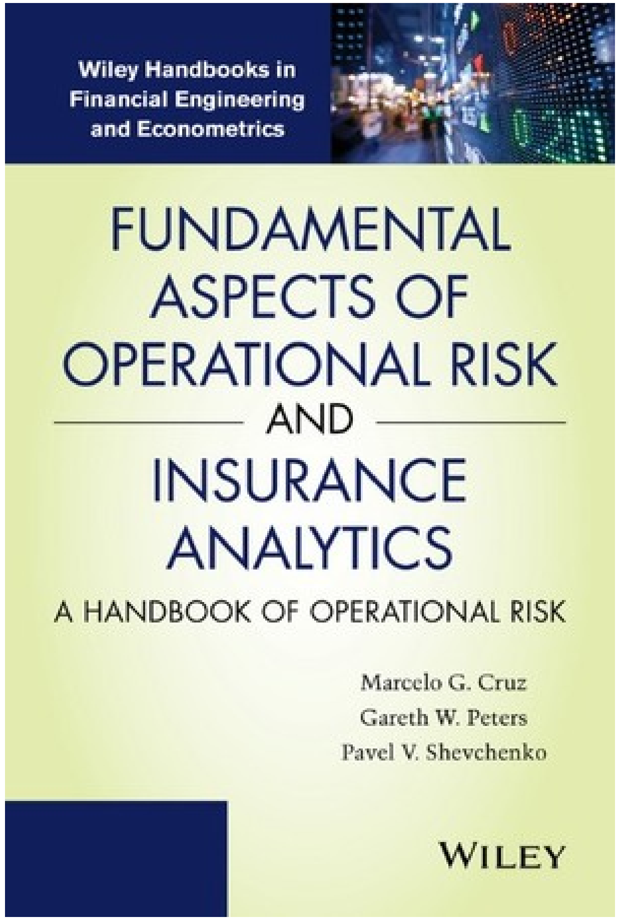
A quote from RISK.net, 13 March 2013:

- "In the past three years, we have seen, again and again, massive legal claims against banks that dwarf the sum of all the other operational risk loss events. That's a major issue, and I don't think many of the current risk models are reflecting this reality," says Paul Embrechts, professor of mathematics at ETH, a university in Zürich.
- He is referring to cases such as those arising from the pre-crisis mortgage boom, which produced a $25 billion settlement in February 2012 between the US and five mortgage servicers: Ally Financial, BAML, Citi, JP Morgan and Wells Fargo. More recent regulatory settlements include December's $1.9 billion money-laundering penalty for HSBC and the $1.5 billion Libor rigging fine for UBS. With US banks' mortgage misdeeds still not fully settled, and regulators around the world still pursuing Libor investigations – while civil cases wait in the wings – the pain is likely to continue.

Quotes from "Bank Capital for Operational Risk: A Tale of Fragility and Instability", M. Ames, T. Schuermann, H.S. Scott, February 10, 2014:

- On May 16, 2012, Thomas Curry, the Comptroller of the Currency (head of the OCC), said in a speech that bank supervisors are seeing "operational risk eclipse credit risk as a safety and soundness challenge." This represents a real departure from the past when concern was primarily focused on credit and market risk. A major component of operational risk is legal liability, and the recent financial crisis, a credit crisis par excellence, has been followed by wave after wave of legal settlements from incidents related to the crisis.
- To again quote Curry (2012), "The risk of operational failure is embedded in every activity and product of an institution."

As a consequence, a lot has been written on the topic:



2015, 900 pages!

etc …

The regulatory approaches towards OpRisk capital :

The Elementary Approaches:

- The Basic Indicator Approach $\mathrm{RC}^t_{\mathrm{BI}}(\mathrm{OR}) = \frac{1}{Z_t} \sum_{i=1}^{3} \alpha \max(\mathrm{GI}^{t-i}, 0)$

risk weight 15%

where $Z_t = \sum_{i=1}^{3} I_{\{\mathrm{GI}^{t-i}>0\}}$ and GI = Gross Income (year t-i)

- The Standardized Approach

$$\mathrm{RC}^t_{\mathrm{S}}(\mathrm{OR}) = \frac{1}{3} \sum_{i=1}^{3} \max\left[\sum_{j=1}^{8} \beta_j \mathrm{GI}_j^{t-i}, 0\right]$$

where the regulatory weight factors 12% ≤ β_j ≤ 18%, j = 1, … , 8 (BLs)
Note: recent BCBS document yields different weights and suggest replacing GI (Gross Income) by a new, so-called Business Indicator (BI).

The Advanced Approaches: AMA and in particular LDA → next slide

The Main LDA-Steps towards a Total Capital Charge

(LDA = Loss Distribution Approach, within AMA = Advanced Measurement Approach)

- Estimation of marginal VaR: $$\boxed{\widehat{\mathrm{VaR}}_\alpha^1, \ldots, \widehat{\mathrm{VaR}}_\alpha^d} \quad (1)$$

(α = p throughout)

- Additional Aggregation: $$\boxed{\widehat{\mathrm{VaR}}_\alpha^+ = \sum_{k=1}^{d} \widehat{\mathrm{VaR}}_\alpha^k} \quad (2)$$

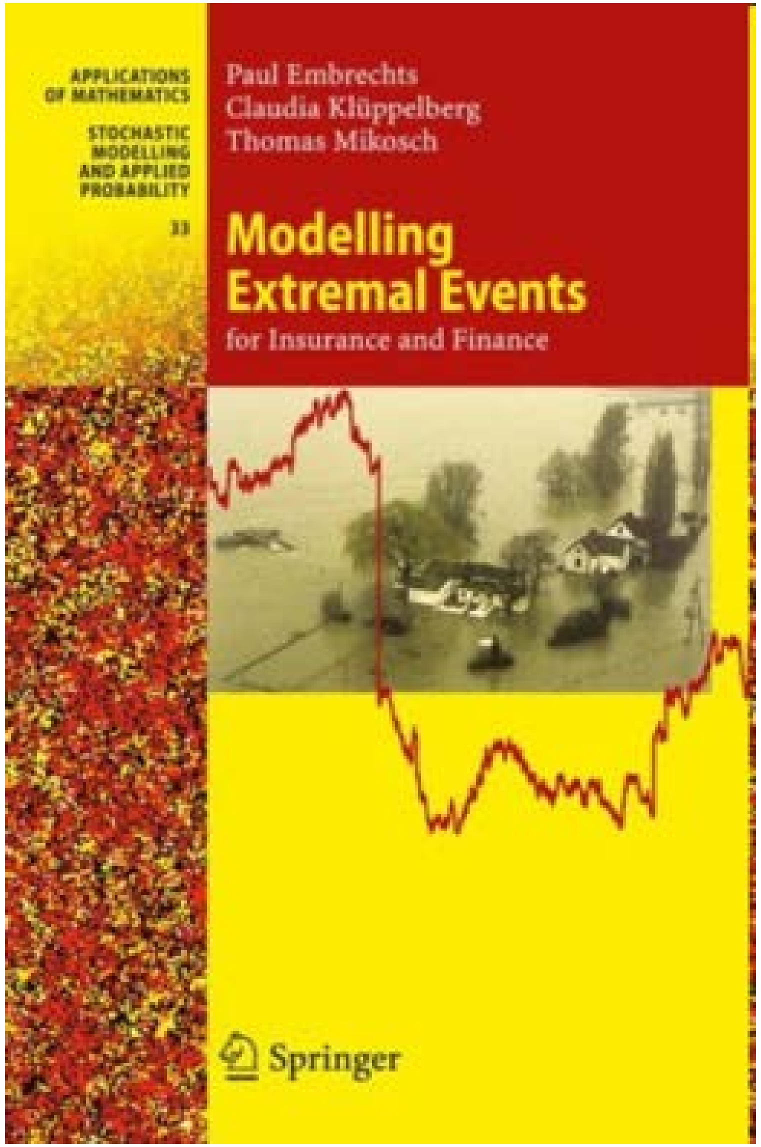
Two very big IFs

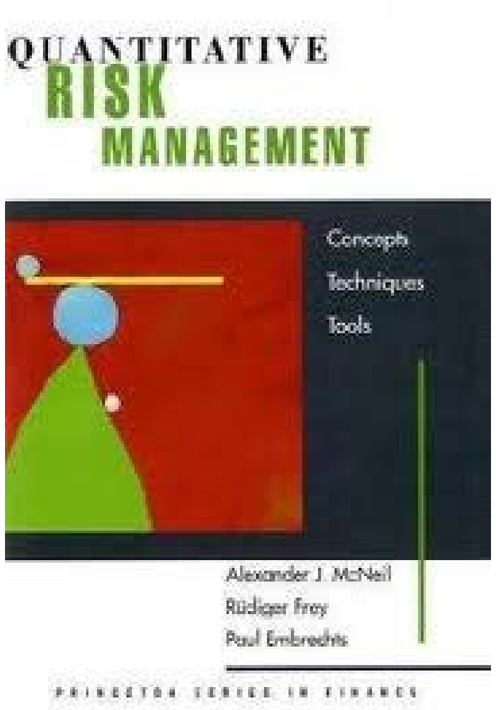
- Diversification: Operational Risk Capital = $$\boxed{\mathrm{VaR}_\alpha^{\mathrm{real}} \overset{?}{<} \widehat{\mathrm{VaR}}_\alpha^+} \quad (3)$$

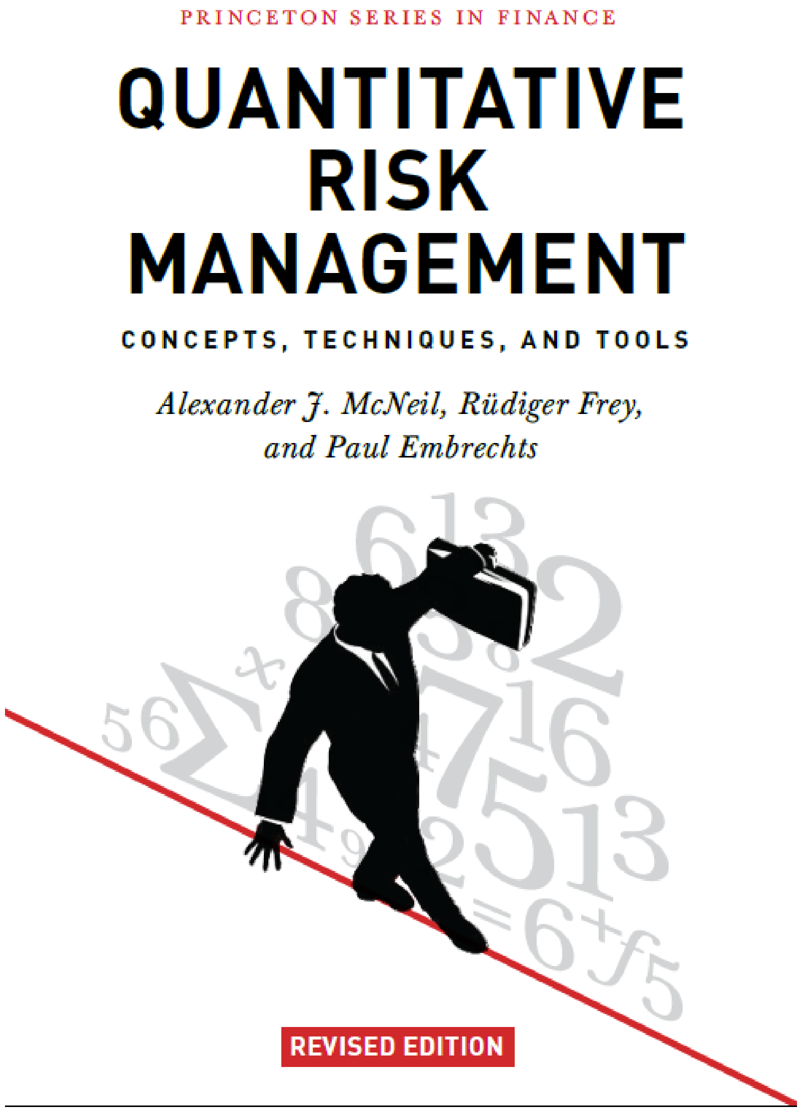
Some comments on (1), (2) and (3)

- For (1), estimating extreme quantiles, an EVT-based picture tells a thousand words → next two slides!
- Equation (2) is fully understood: Given that d risks are comonotone, then the VaR of their sum is the sum of their VaRs, hence (2) yields the VaR of the aggregate position under comonotonicity ("maximal correlation, perfect positive dependence, … ")
- Definition: Random variables $X_1, \dots, X_d$ are comonotone if there exists a random variable Z and d increasing functions $\xi_1, \dots, \xi_d$ so that $X_i = \xi_i(Z)$, almost surely, i=1, … , d.
- For (3): model - and dependence uncertainty (← this talk)

(1) Estimating extreme quantiles (VaR)





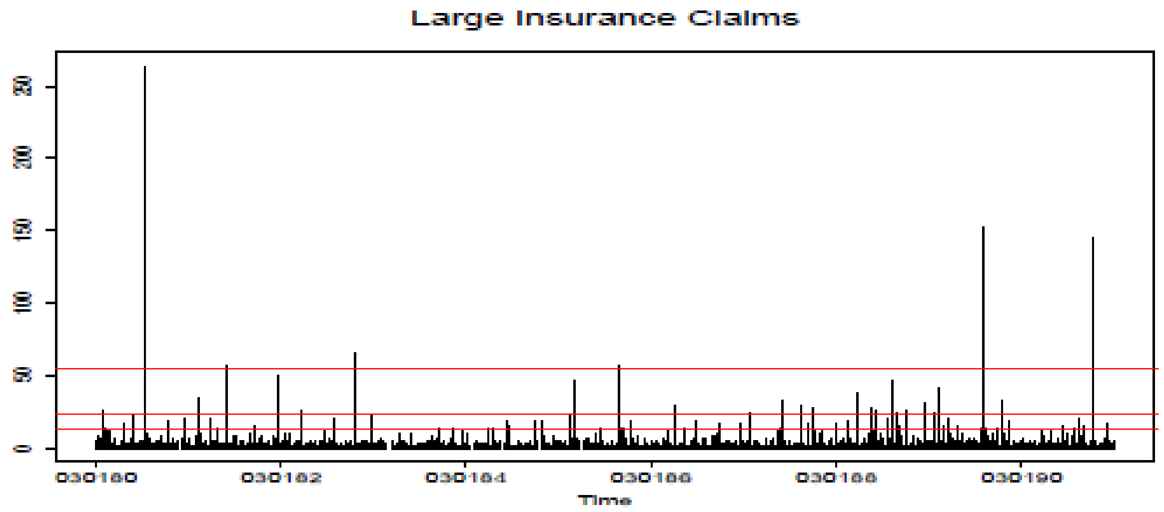


Danish Fire Loss Example

The Danish data consist of 2167 losses exceeding one million Danish Krone from the years 1980 to 1990. The loss figure is a total loss for the event concerned and includes damage to buildings, damage to contents of buildings as well as loss of profits. The data have been adjusted for inflation to reflect 1985 values.

Very similar to OpRisk data!

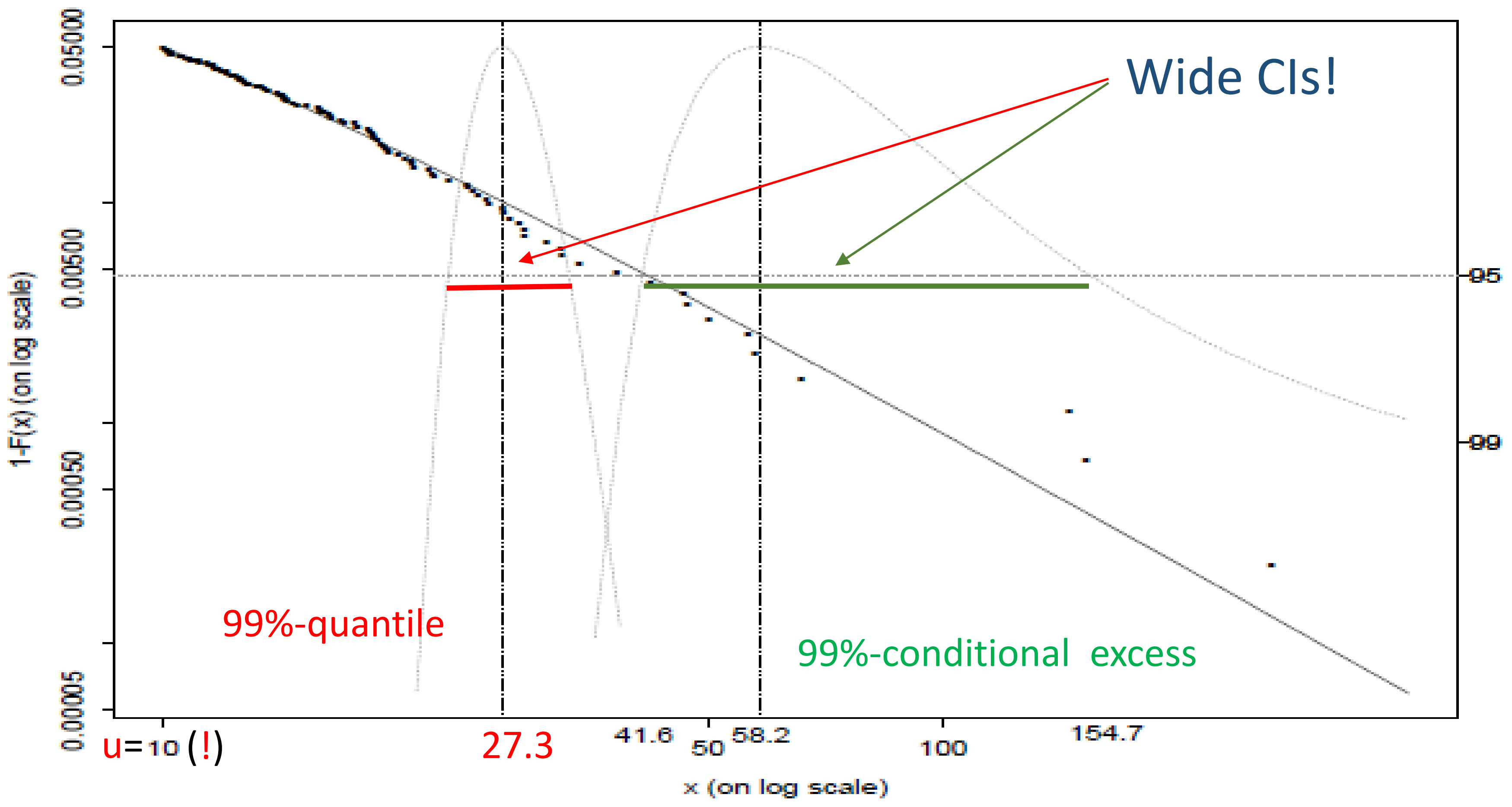
POT u



99%-quantile with 95% CI (Profile Likelihood):

27.3 (23.3, 33.1)

99% Conditional Excess: E(X I X > 27.3) with CI



Concerning (2), recall that

- In general, VaR is not sub-additive, typical such cases occur for risks which are either very heavy-tailed (infinite mean), very skewed or (whatever the marginal dfs, e.g. N(0,1) or Exp(1)) have special dependence: all these cases are relevant for OpRisk!
- VaR is sub-additive (coherent) for multivariate elliptical risk factors.
- VaR and (hence also) ES are additive for comonotonic risks.
- Hence for ES, adding up the ES-contributions from the marginal risk factors always yields an upper bound for ES of the sum, and the upper bound is reached in the comonotonic case.
- For VaR this is NOT TRUE and this is relevant within the OpRisk context!

(3) Model - and Dependence Uncertainty

- Standard Basel II(+) procedure: aggregate the OpRisk losses BL-wise
- Estimate the resulting (8) VaRs
- Add these numbers up leading to a global estimate VaR^+
- Recall de notion and importance of comonotonic dependence
- Invoke a diversification argument to bring down regulatory capital from VaR^+ to a factor (1 – δ) VaR^+ where often δ ≈ 0.3
- However the non-convexity of VaR as a Risk measure may lead to true measures of risk (capital) larger than VaR^+, hence an important question concerns the problem of calculating best-worst bounds on risk measures of portfolio positions in general and VaR and ES more in particular

A fundamental problem in Quantitative Risk Management:

- Risk factors: $\mathbf{X} = (X_1, \ldots, X_d)$
- Model assumption: $X_i \sim F_i$, F_i known, $i = 1, \ldots, d$
- A financial position $\Psi(\mathbf{X})$
- A risk measure/pricing function: $\rho(\Psi(\mathbf{X}))$

Calculate $\rho(\Psi(\mathbf{X}))$

also denoted by S_d

Example:

- $\Psi(\mathbf{X}) = \sum_{i=1}^{d} X_i$
- $\rho = \mathrm{VaR}_p$ or $\rho = \mathrm{ES}_p$

Challenge:

- We need a *joint* model for the random vector $\mathbf{X}$
- Joint models are hard to get by

We will focus on the above special choices of Ψ and ρ.

For a given risk measure ρ denote

$$\overline{\rho}(S_d) = \sup \{ \rho(\sum_{i=1}^{d} X_i) : X_i \sim F_i, i = 1, \ldots, d\}$$

and similarly

$$\underline{\rho}(S_d) = \inf \{ \rho(\sum_{i=1}^{d} X_i) : X_i \sim F_i, i = 1, \ldots, d\}$$

where sup/inf are taken over all joint distribution models for the random vector $(X_1, \ldots, X_d)$ with given marginal dfs $(F_1, \ldots, F_d)$, or equivalently over all d-dimensional copulas.

We will consider as special cases the construction of the ranges:

$$(\underline{\mathrm{VaR}}, \overline{\mathrm{VaR}}) \text{ and } (\underline{\mathrm{ES}}, \overline{\mathrm{ES}})$$

known: comonotonic case

referred to as dependence-uncertainty ranges.

Summary of existing results:

$d = 2$:

- fully solved analytically

$d \geq 3$:

- Homogeneous model ($F_1 = \cdots = F_d$)
    - $\underline{\mathrm{ES}}_p(S_d)$ solved analytically for decreasing densities, e.g. Pareto, Exponential
    - $\overline{\mathrm{VaR}}_p(S_d)$ solved analytically for tail-decreasing densities, e.g. Pareto, Gamma, Log-normal
- Inhomogeneous model
    - Few analytical results: current research
- Numerical methods available: Rearrangement Algorithm

Sharp(!) bounds in the homogeneous case:

Sharp VaR bounds (Wang, Peng and Yang, 2013)

Suppose that the density function of F is decreasing on $[b, \infty)$ for some $b \in \mathbb{R}$. Then, for $p \in [F(b), 1)$, and $X \overset{d}{\sim} F$,

$$\overline{\mathrm{VaR}}_p(S_d) = d\mathbb{E}[X|X \in [F^{-1}(p + (d-1)c), F^{-1}(1-c)]],$$

where c is the smallest number in $[0, \frac{1}{d}(1-p)]$ such that

$$\int_{p+(d-1)c}^{1-c} F^{-1}(t)\mathrm{d}t \geq \frac{1-p-dc}{d}((d-1)F^{-1}(p+(d-1)c) + F^{-1}(1-c)).$$

Condition!

Red part clearly has an ES-type form.

- $c = 0$: $\overline{\mathrm{VaR}}_p(S_d) = \overline{\mathrm{ES}}_p(S_d)$.

More general result in the background!

Sharp VaR bounds II

Suppose that the density function of F is decreasing on its support. Then for $p \in (0,1)$ and $X \overset{d}{\sim} F$,

$$\underline{\mathrm{VaR}}_p(S_d) = \max\{(d-1)F^{-1}(0) + F^{-1}(p), d\mathbb{E}[X|X \le F^{-1}(p)]\}.$$

Stronger condition!

Sharp ES bounds (Bernard, Jiang and Wang, 2014)

Suppose that the density function of F is decreasing on its support. Then for $p \in (1-dc, 1)$, $q = (1-p)/d$ and $X \overset{d}{\sim} F$,

$$\underline{\mathrm{ES}}_p(S_d) = \frac{1}{q}\int_0^q \left((d-1)F^{-1}((d-1)t) + F^{-1}(1-t)\right) \mathrm{d}t,$$
$$= (d-1)^2 \mathrm{LES}_{(d-1)q}(X) + \mathrm{ES}_{1-q}(X),$$

Left-tail-ES

where c is the smallest number in $[0, \frac{1}{d}]$ such that

$$\int_{(d-1)c}^{1-c} F^{-1}(t)\mathrm{d}t \ge \frac{1-dc}{d}((d-1)F^{-1}((d-1)c) + F^{-1}(1-c)).$$

- One large outcome is coupled with $d-1$ small outcomes. $\longleftarrow$: basic idea behind the proof

Bounds in the inhomogeneous case: RA

Rearrangement Algorithm (RA): Embrecths, Puccetti and Rüschendorf (2013).

- A fast numerical procedure
- Based on the CM-idea
- Discretization of relevant quantile regions
- d possibly large
- Applicable to $\overline{\mathrm{VaR}}_p$, $\underline{\mathrm{VaR}}_p$ and $\underline{\mathrm{ES}}_p$

CM = Complete Mixability

Complete mixability, Wang and Wang (2011)

A distribution function F on $\mathbb{R}$ is called d-completely mixable (d-CM) if there exist d random variables $X_1, \ldots, X_d \sim F$ such that

$$\mathbb{P}(X_1 + \cdots + X_d = dk) = 1,$$

for some $k \in \mathbb{R}$.

Related concepts:

- d-mixability
- inhomogeneous case
- strong negative dependence
- general extremal dependence, …

Home Sample codes R packages Further Extensions OR framework Contributors License

The Rearrangement Algorithm project

The Rearrangement Algorithm (RA) is an algorithm which has been introduced in [1] to compute numerically sharp lower and upper bounds on the distribution of a function of a number of dependent random variables having fixed marginal distributions.

The algorithm has been then developed further to:

- compute sharp bounds for the VaR/ES of high-dimensional portfolios having fixed marginal distributions; see [2], [3].
- compute sharp lower and upper bounds on the expected value of a supermodular function of d random variables having fixed marginal distributions; see [4].

For full details, see https://sites.google.com/site/rearrangementalgorithm/

Example 1: $P(X_i > x) = (1+x)^{-2}$, $x \geq 0$, i = 1 , … , d

Bounds on VaR and ES for the sum of d Pareto(2) distributed rvs for $p = 0.999$; VaR_p^+ corresponds to the comonotonic case.

DU-gaps

434

220

can be explained

	$d = 8$	$d = 56$
$\underline{\mathrm{VaR}}_p$	31	53
$\underline{\mathrm{ES}}_p$	178	472
VaR_p^+	245	1715
$\overline{\mathrm{VaR}}_p$	465	3454
$\overline{\mathrm{ES}}_p$	498	3486
$\overline{\mathrm{VaR}}_p/\mathrm{VaR}_p^+$	1.898	2.014
$\overline{\mathrm{ES}}_p/\overline{\mathrm{VaR}}_p$	1.071	1.009

Comonotonic case: sum of marginal VaRs = d x marginal VaR

Comonotonic case: sum of marginal ESs = d x marginal ES

+/- factor 2 can be explained: Karamata's Theorem

+/- factor 1 can be explained : next slide

Two theorems (Embrechts, Wang, Wang, 2104):

Theorem 1:

Suppose the continuous distributions F_i, $i \in \mathbb{N}$ satisfy that for $X_i \sim F_i$ and some $p \in (0,1)$,

(i) $\mathbb{E}[|X_i - \mathbb{E}[X_i]|^k]$ *is uniformly bounded for some* $k > 1$;

(ii) $\liminf_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \mathrm{ES}_p(X_i) > 0.$

Then as $d \to \infty$,

$$\frac{\overline{\mathrm{ES}}_p(S_d)}{\overline{\mathrm{VaR}}_p(S_d)} = 1 + O(d^{1/k-1}).$$

Theorem 2:

Take $1 > q \geq p > 0$. *Suppose that the continuous distributions* F_i, $i \in \mathbb{N}$, *satisfy (i) and (iii), and* $\limsup_{d\to\infty} \frac{\sum_{i=1}^{d}\mathbb{E}[X_i]}{\sum_{i=1}^{d}\mathrm{ES}_p(X_i)} < 1$, *then*

$$\liminf_{d\to\infty} \frac{\overline{\mathrm{VaR}}_q(S_d) - \underline{\mathrm{VaR}}_q(S_d)}{\overline{\mathrm{ES}}_p(S_d) - \underline{\mathrm{ES}}_p(S_d)} \geq 1.$$

A second example (inhomogeneous case):

ES and VaR of $S_d = X_1 + \cdots + X_d$, where

- $X_i \sim \text{Pareto}(2 + 0.1i)$, $i = 1, \dots, 5$;
- $X_i \sim \text{Exp}(i - 5)$, $i = 6, \dots, 10$;
- $X_i \sim \text{Log–Normal}(0, (0.1(i - 10))^2)$, $i = 11, \dots, 20$.

	$d = 5$			$d = 20$		
	best	worst	spread	best	worst	spread
$\text{ES}_{0.975}$	22.48	44.88	22.40	29.15	102.35	73.20
$\text{VaR}_{0.975}$	9.79	41.46	31.67	21.44	100.65	79.21
$\text{VaR}_{0.9875}$	12.06	56.21	44.16	22.12	126.63	104.51
$\text{VaR}_{0.99}$	12.96	62.01	49.05	22.29	136.30	114.01
$\frac{\overline{\text{ES}}_{0.975}}{\overline{\text{VaR}}_{0.975}}$		1.08			1.02	

Conclusions

- Operational Risk is a very important risk class, but defies reliable quantitative modelling
- More standardisation within the AMA/LDA is called for, do not allow for full modelling freedom: danger of backwards engineering
- Use lower confidence levels together with regulatory defined scaling
- Split legal risk from other Operational Risk classes and decide on separate treatment
- Make data available for scientific research
- Operational Risk type of data may lead to interesting statistical research questions which are relevant in a wider context, like (*) →

Some (extra) references:

- P. Embrechts, B. Wang, R. Wang (2014) Aggregation robustness and model uncertainty of regulatory risk measures. Finance and Stochastics, to appear (2015)
- (*) V. Chavez-Demoulin, P. Embrechts, M. Hofert (2014) An extreme value approach for modeling Operational Risk losses depending on covariates. Journal of Risk and Insurance, to appear (2015)

See www.math.ethz.ch/~embrechts for details.

Thank You!