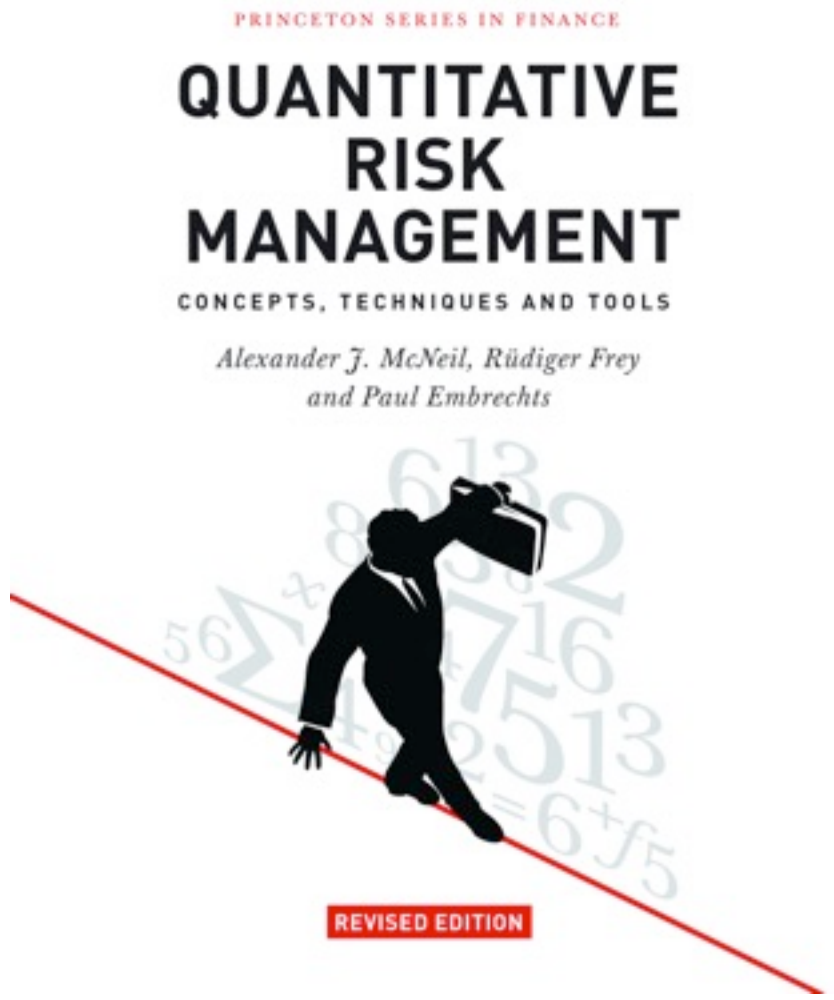
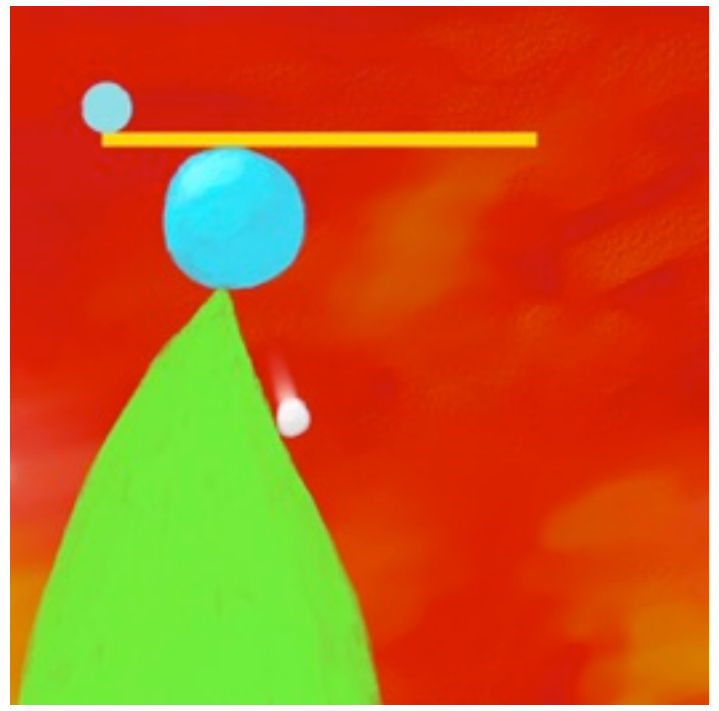
QUANTITATIVE RISK MANAGEMENT IN R

Welcome to the course!

About me

- Professor in mathematical statistics, actuarial science, and quantitative finance
- Author of *Quantitative Risk Management: Concepts, Techniques & Tools* with R. Frey and P. Embrechts
- Creator of qrmtutorial.org with M. Hofert
- Contributor to R packages including `qrmdata` and `qrmtools`





The objective of QRM

- In **quantitative risk management** (QRM), we quantify the risk of a portfolio
- Measuring risk is first step towards managing risk
- Managing risk:
  - Selling assets, diversifying portfolios, implementing hedging with derivatives
  - Maintaining sufficient capital to withstand losses
- **Value-at-risk** (VaR) is a well-known measure of risk

Risk factors

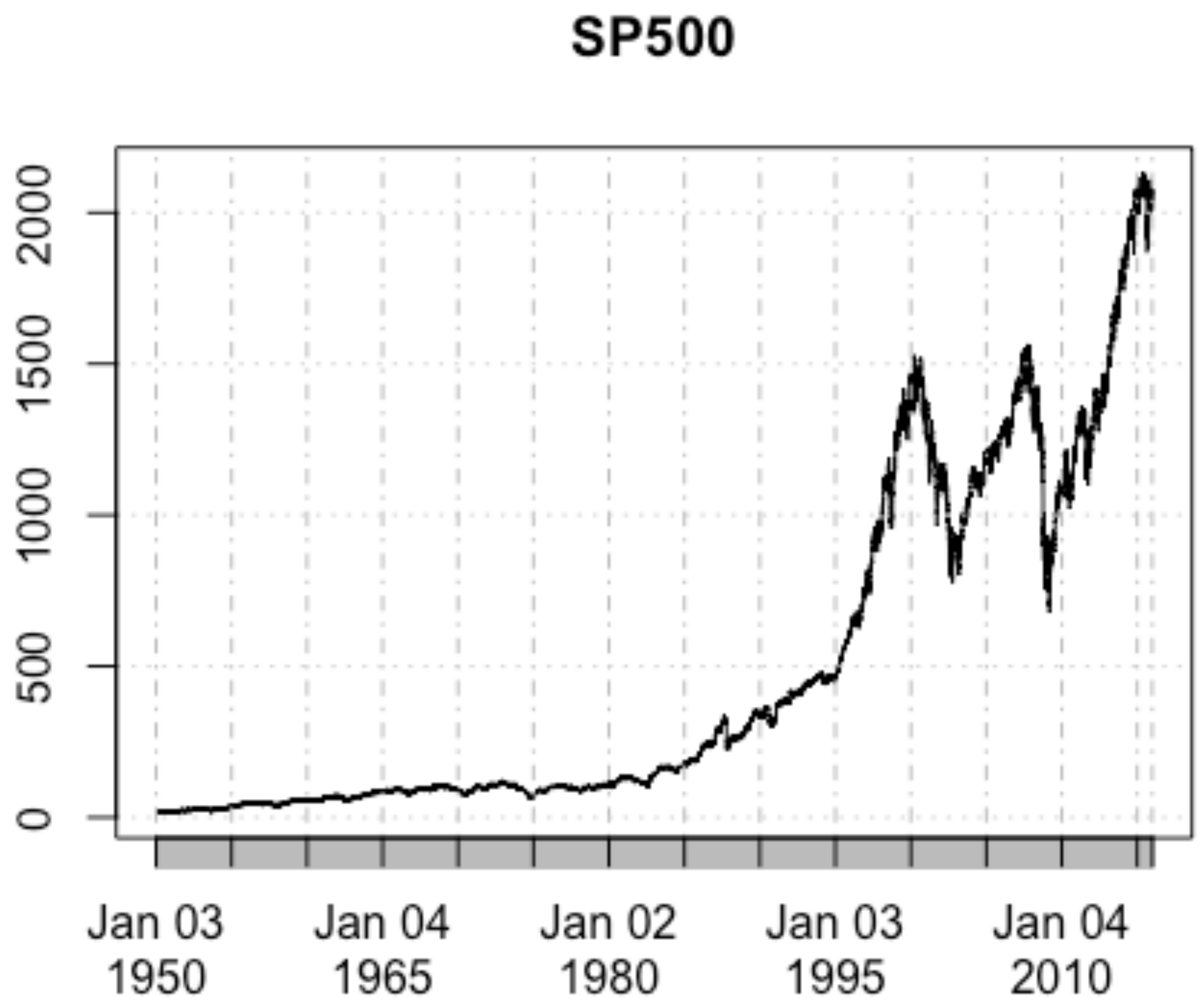
- Value of a portfolio depends on many **risk factors**
- Examples: equity indexes/prices, FX rates, interest rates
- Let's look at the S&P 500 index

Analyzing risk factors with R

```
> library(qrmdata)

> data(SP500)

> head(SP500,  n = 3)
                   ^GSPC
1950-01-03 16.66
1950-01-04 16.85
1950-01-05 16.93

> tail(SP500, n = 3)
                      ^GSPC
2015-12-29 2078.36
2015-12-30 2063.36
2015-12-31 2043.94
```

Plotting risk factors

```
> plot(SP500)
```



Let's practice!

Risk-factor returns

Risk-factor returns

- Changes in risk factors are **risk-factor returns** or **returns**
- Let (Z_t) denote a time series of risk factor values
- Common definitions of returns (X_t) :

$$X_t = Z_t - Z_{t-1} \qquad \text{(simple returns)}$$

$$X_t = \frac{Z_t - Z_{t-1}}{Z_{t-1}} \qquad \text{(relative returns)}$$

  - 0.02 = 2% gain, -0.03 = 3% loss

$$X_t = \ln(Z_t) - \ln(Z_{t-1}) \qquad \text{(log-returns)}$$

Properties of log-returns

- Resulting risk factors cannot become negative
- Very close to relative returns for small changes:

$$\ln(Z_t) - \ln(Z_{t-1}) \approx \frac{Z_t - Z_{t-1}}{Z_{t-1}}$$

- Easy to aggregate by summation to obtain longer-interval log-returns
- Independent normal if risk factors follow **geometric Brownian motion** (GBM)

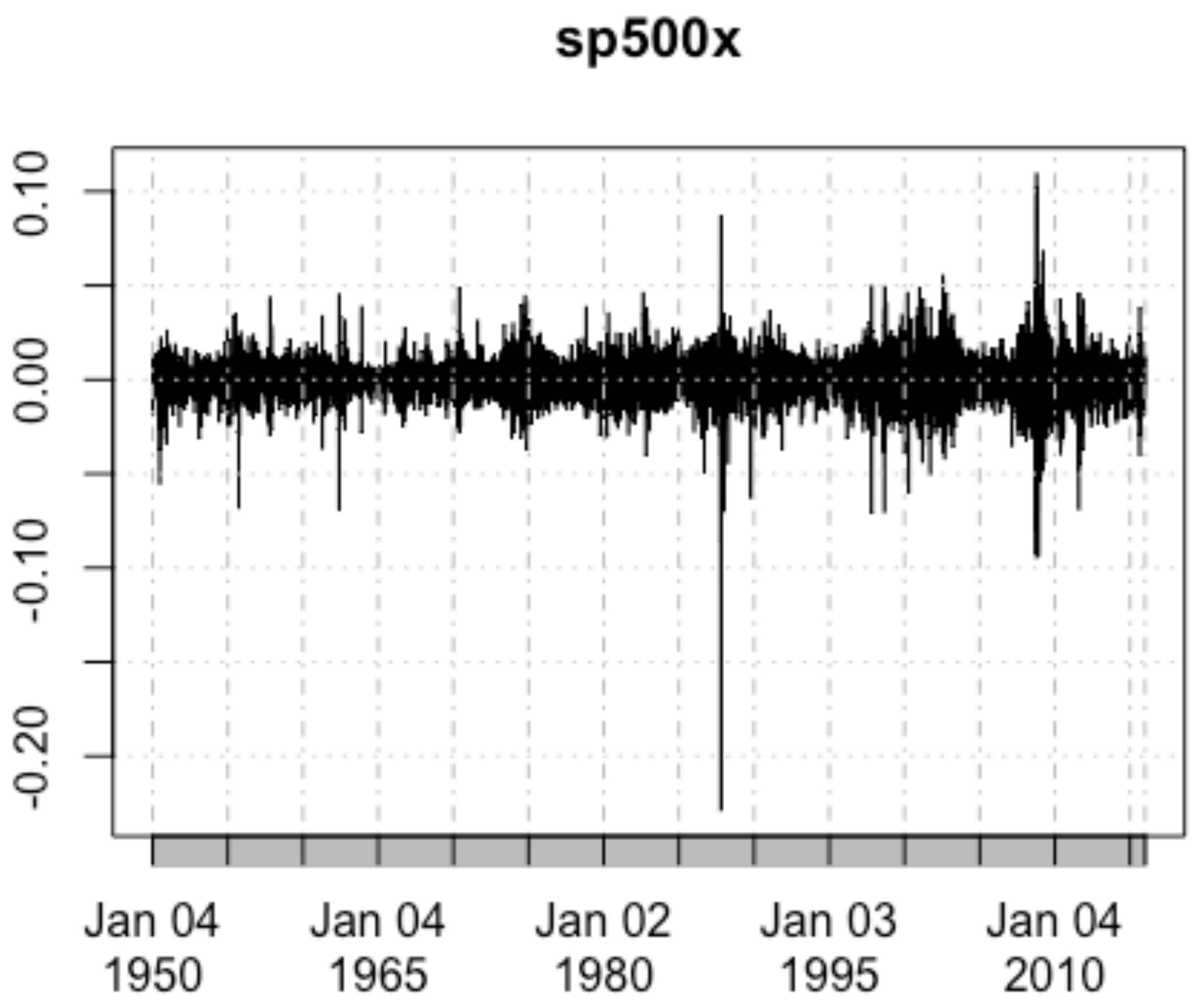
Log-returns in R

```
> sp500x <- diff(log(SP500))
> head(sp500x, n = 3)  # note the NA in first position
                           ^GSPC
1950-01-03          NA
1950-01-04 0.011340020
1950-01-05 0.004736539
```

```
> sp500x <- diff(log(SP500))[-1]
> head(sp500x)
                            ^GSPC
1950-01-04  0.011340020
1950-01-05  0.004736539
1950-01-06  0.002948985
1950-01-09  0.005872007
1950-01-10 -0.002931635
1950-01-11  0.003516944
```

Log-returns in R (2)

```
> plot(sp500x)
```



Let's practice!

Aggregating log-returns

Aggregating log-returns

- Just add them up!
- Assume (X_t) are daily log-returns calculated from risk-factor values (Z_t)
- Log-returns for a trading week is the sum of log-returns for each trading day:

$$\ln(Z_{t+5}) - \ln(Z_t) = \sum_{i=1}^{5} X_{t+i}$$

- Similar for other time horizons

Aggregating log-returns in R

- Use the `sum()` function within `apply.weekly()` and `apply.monthly()` in the `xts` package

```
> sp500x_w <- apply.weekly(sp500x, sum)
> head(sp500x_w, n = 3)
                           ^GSPC
1950-01-09  0.02489755
1950-01-16 -0.02130264
1950-01-23  0.01189081
```

```
> sp500x_m <- apply.monthly(sp500x, sum)
> head(sp500x_m, n = 3)
                           ^GSPC
1950-01-31 0.023139508
1950-02-28 0.009921296
1950-03-31 0.004056917
```

Let's practice!

Exploring other kinds of risk factors

Exploring other kinds of risk factors

- So far we have looked at:
  - Calculating log-returns and aggregating log-returns over longer intervals
  - Equity data, indexes and single stocks, and **foreign-exchange** (FX) data
- Two other categories of risk factors:
  - Commodities prices
  - Yields of zero-coupon bonds

Commodities data and interest-rate data

- Commodities such as gold and oil prices
  - Do log-returns behave like stocks?
- Government bonds - value depends on interest rates
  - Consider **yields of zero-coupon bonds** as risk factors

Bond prices

- Let $p(t, T)$ denote the price at time small t of a zero-coupon bond paying one unit at maturity T
- $p(0, 10)$: price at $t = 0$ of bond maturing at $T = 10$
- $p(0, 5)$: price at $t = 0$ of bond maturing at $T = 5$
- $p(5, 10)$: price at $t = 5$ of bond maturing at $T = 10$

Yields as risk factors

- The **yield** y(t, T) is defined by the equation:

$$y(t,T) = \frac{-\ln p(t,T)}{T-t}$$

- y(t, 10): yield for a 10-year bond acquired at time t
- y(t, 5): yield for a 5-year bond acquired at time t
- Advantage of yields: comparable across maturities T
- The mapping T to y(t, T) is yield curve at time t
- Log-returns or simple returns of yields?

Let's practice!

The normal distribution

Definition of normal

- If risk factors follow GBM, then log-returns should be independent normal
- Is this the case?
- A variable x is normal if it has density:

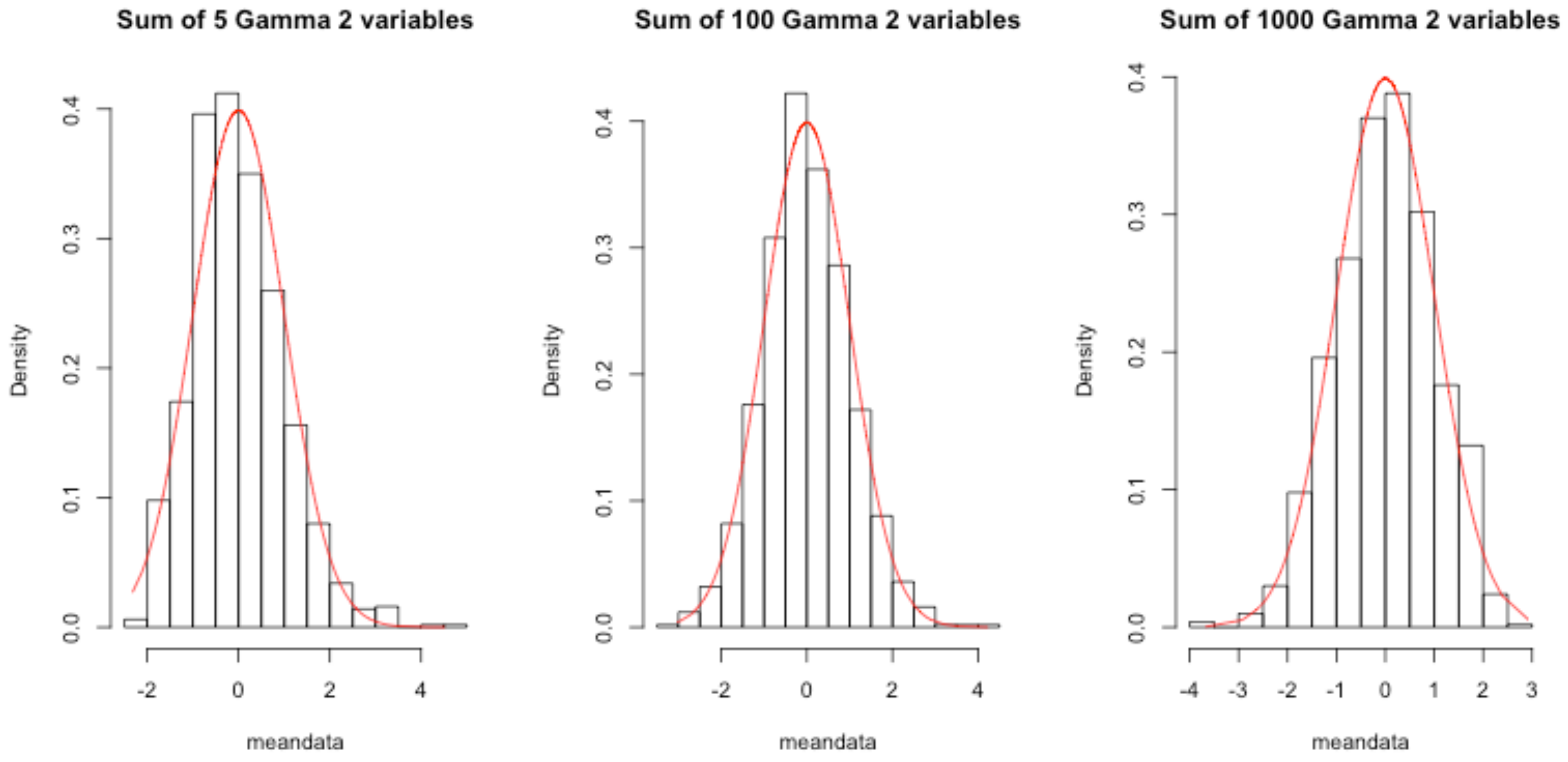
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}}$$

Depends on two parameters: μ and σ

Properties of the normal

- μ is the mean and σ^2 is the variance
- Usual notation: $X \sim N(\mu, \sigma^2)$
- Parameters easily estimated from data
- Sum of 2+ independent normal variables is also normal

Central limit theorem (CLT)



How to estimate a normal distribution

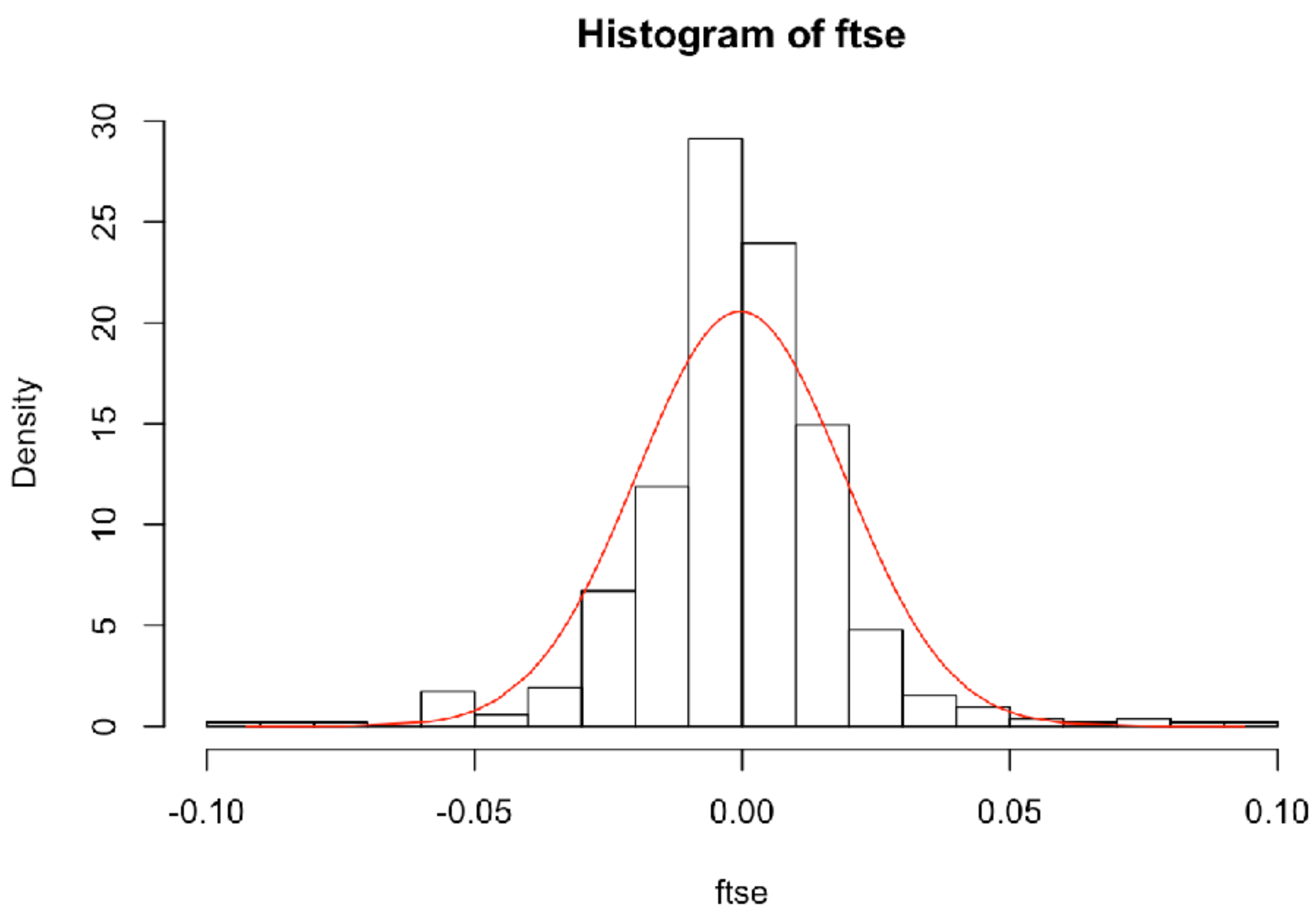
- Data: $X_1, \ldots, X_n$
- **Method of moments:** $$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} X_t$$
  $$\hat{\sigma}_u^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \hat{\mu})^2$$
- Application to FTSE log-returns from 2008-09

FTSE example

```
> head(ftse)
[1] -0.09264548 -0.08178433 -0.07428657 -0.05870079 -0.05637430
-0.05496918

> tail(ftse)
[1] 0.05266208 0.06006960 0.07742977 0.07936751 0.08469137
0.09384244

> mu <- mean(ftse)
> sigma <- sd(ftse)
> c(mu, sigma)
[1] -0.0003378627  0.0194090385
```

Displaying the fitted normal

```
> hist(ftse, nclass = 20, probability = TRUE)
> lines(ftse, dnorm(ftse, mean = mu, sd = sigma), col = "red")
```



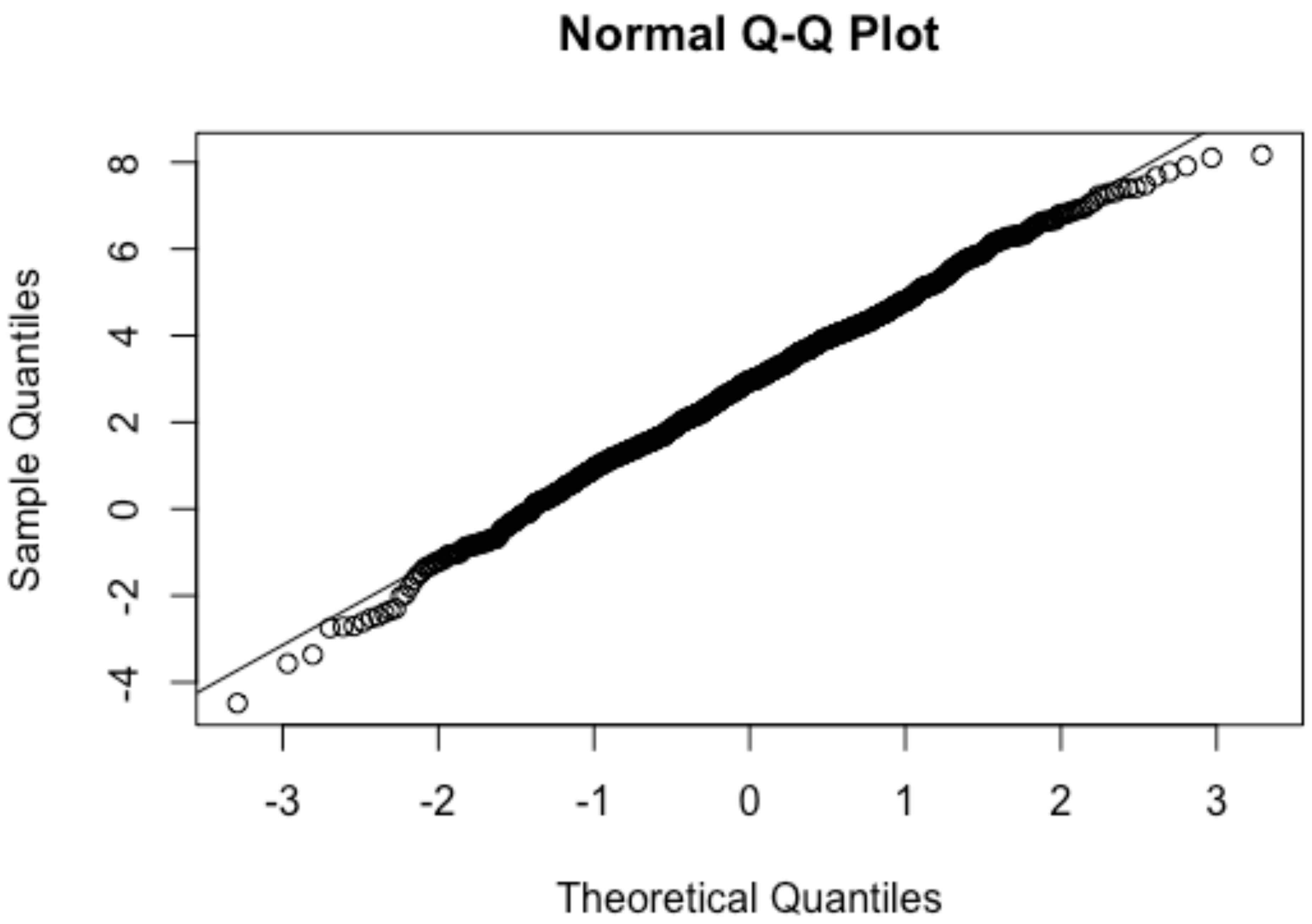
Let’s practice!

Testing for normality

How to test for normality

- Use the **quantile-quantile plot** (Q-Q plot)
- Sample quantiles of data versus theoretical quantiles of a normal distribution

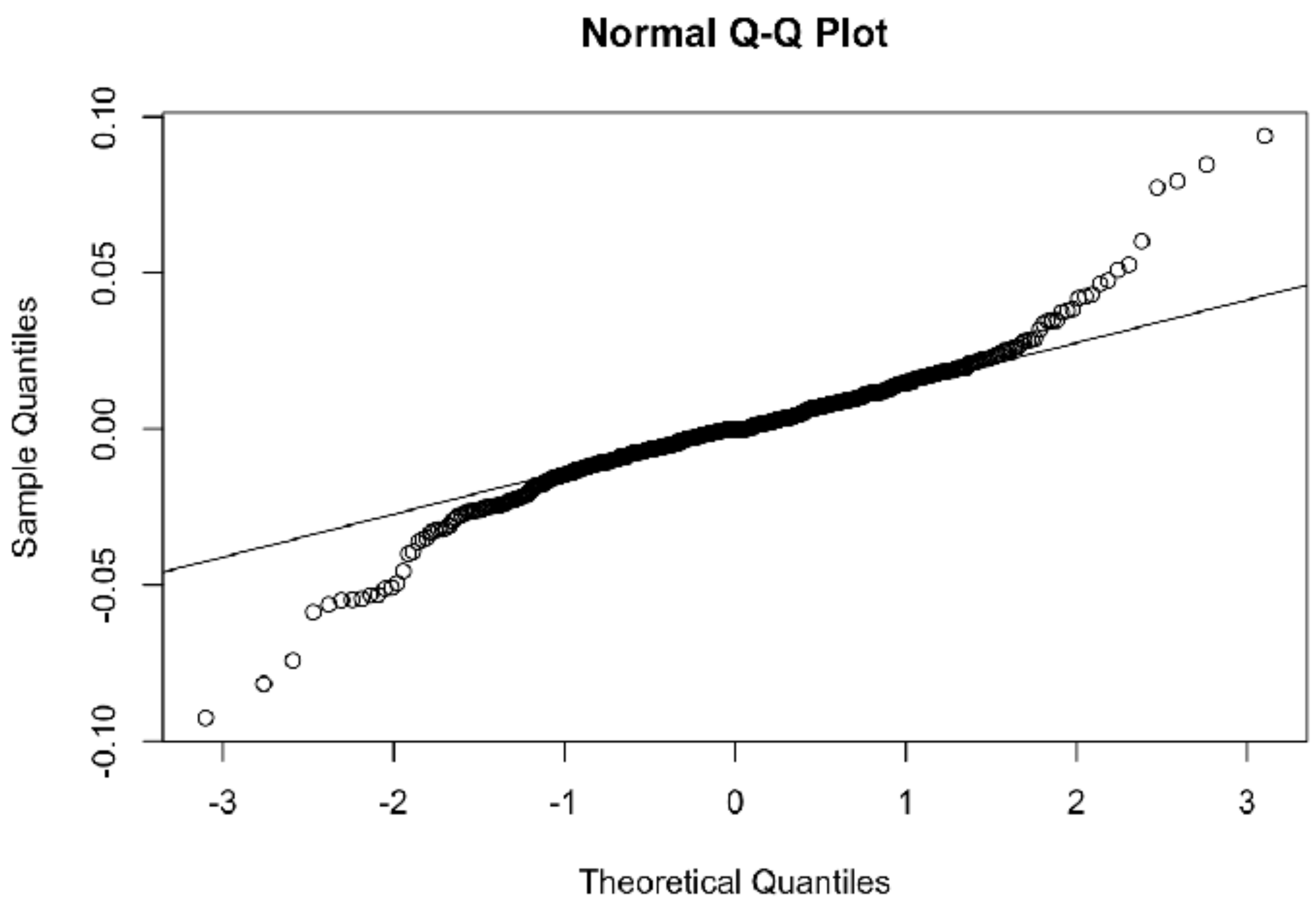
```
> data <- rnorm(1000,
                mean = 3,
                sd = 2)
> qqnorm(data)
> qqline(data)
```



Interpreting the Q-Q plot

- Data with heavier tails than normal: inverted S shape
- Data with lighter tails than normal: S shape
- Data from a very skewed distribution: curved shape

```
> qqnorm(ftse)
> qqline(ftse)
```

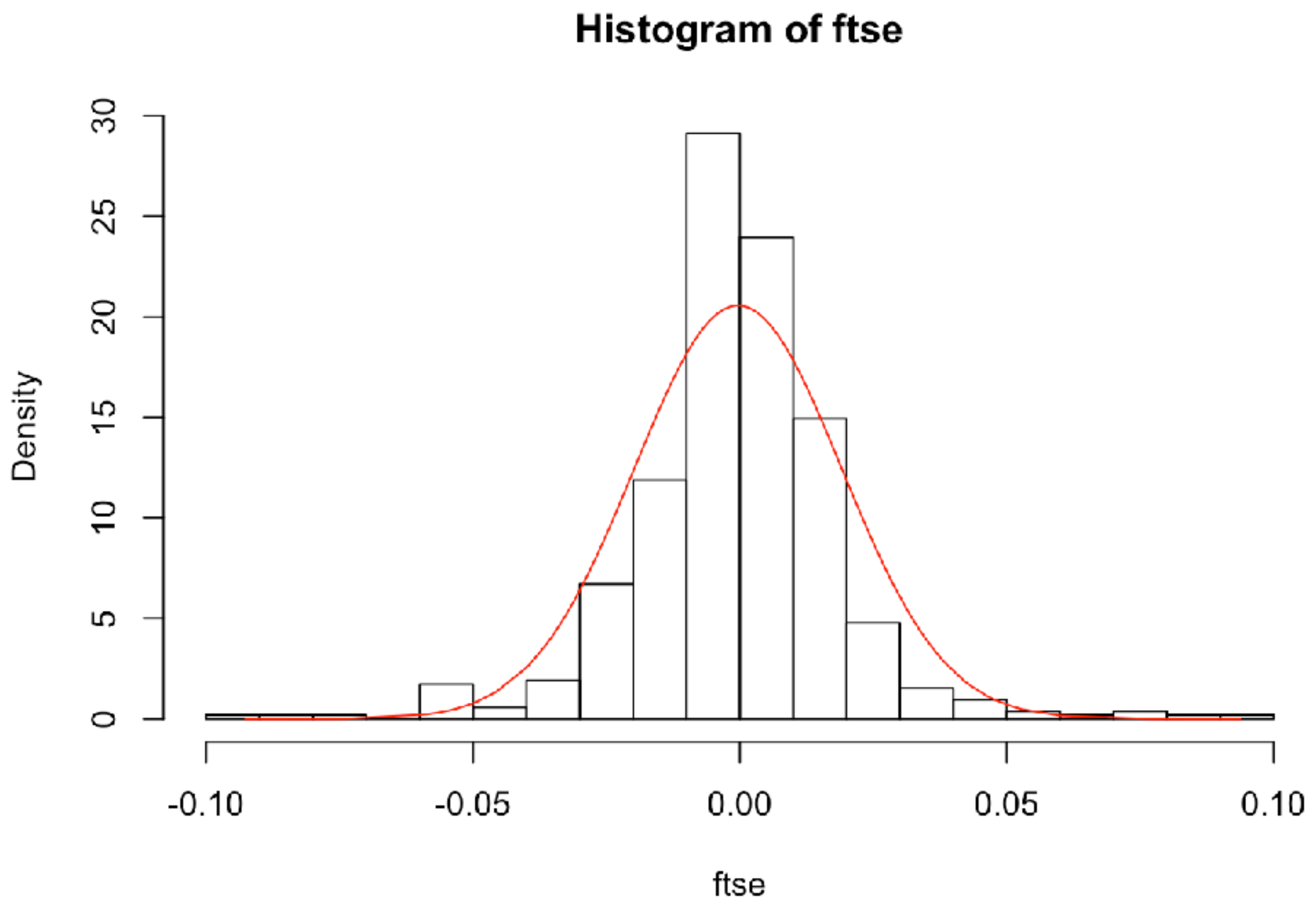


Let’s practice!

Skewness, kurtosis and the Jarque-Bera test

Skewness and kurtosis

- **Skewness** (b) is a measure of asymmetry
- **Kurtosis** (k) is a measure of heavy-tailedness
- Skewness and kurtosis of normal are 0 and 3, respectively

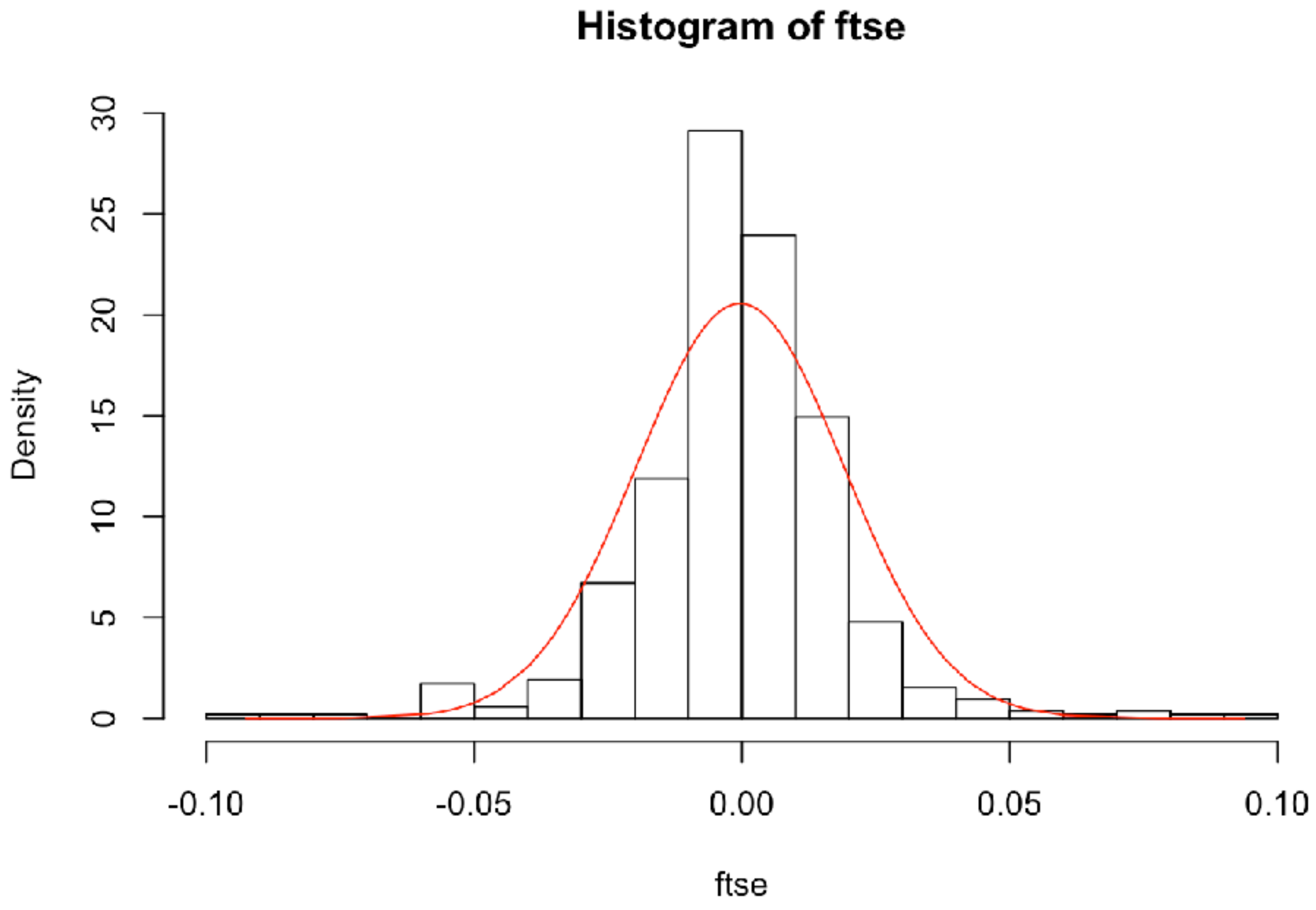


$$= \frac{1}{n}\frac{\sum_{t=1}^{n}(X_t - \hat{\mu})^3}{\hat{\sigma}^3}$$

$$= \frac{1}{n}\frac{\sum_{t=1}^{n}(X_t - \hat{\mu})^4}{\hat{\sigma}^4}$$

Skewness and kurtosis (II)

```
> library(moments)

> skewness(ftse)
[1] -0.01187921

> kurtosis(ftse)
[1] 7.437121
```



The Jarque-Bera test

- Compares skewness and kurtosis of data with theoretical normal values (0 and 3)
- Detects skewness, heavy tails, or both

$$T = \frac{1}{6}n\left(b^2 + \frac{1}{4}(k-3)^2\right)$$

```
> jarque.test(ftse)

  Jarque-Bera Normality Test

data:  ftse
JB = 428.23, p-value < 2.2e-16
alternative hypothesis: greater
```

Longer-interval and overlapping returns

- Daily returns are usually very non-normal
- What about longer-intervals returns?
- Weekly, monthly, quarterly returns obtained by summation
- Recall CLT - suggests they may be more normal
- Reduce quantity of data so tests are weaker
- Can also analyze **overlapping** or **moving sums** of returns

Let's practice!

The Student t distribution

The Student t distribution

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\frac{\nu+1}{2}}$$

- This distribution has three parameters: μ, σ, ν
- Small values of ν give heavier tails
- As ν gets larger the distribution tends to normal

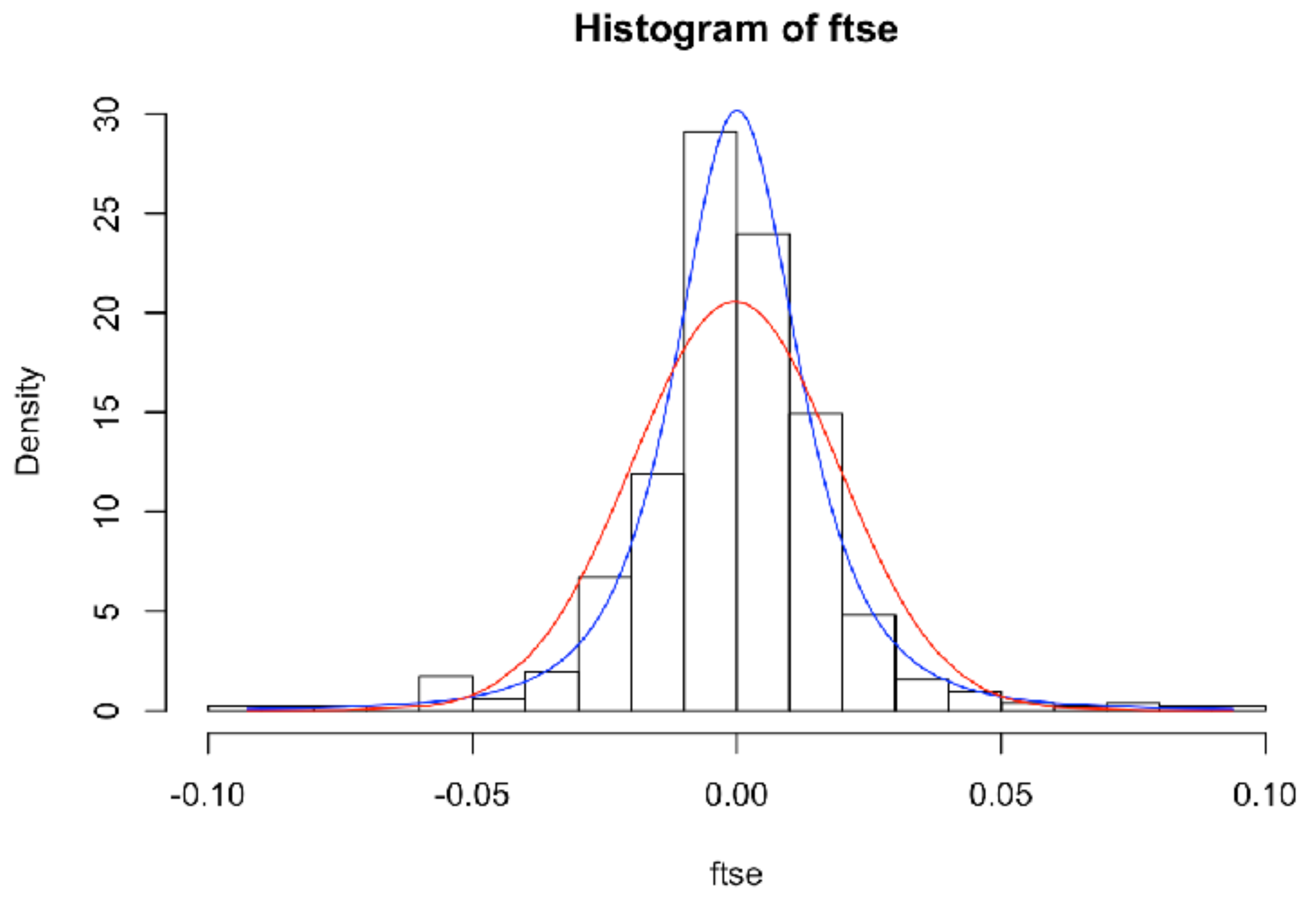
Fitting the Student t distribution

- Method of **maximum likelihood** (ML)
- `fit.st()` in QRM package
- Small ν value (2.95) for FTSE log-returns from 2008-09

```
> library(QRM)

> tfit <- fit.st(ftse)
> tpars <- tfit$par.ests
> tpars
          nu           mu        sigma
2.949514e+00 4.429863e-05 1.216422e-02

> nu <- tpars[1]
> mu <- tpars[2]
> sigma <- tpars[3]
```

Displaying the fitted Student t distribution

```
> hist(ftse, nclass = 20, probability = TRUE)
> lines(ftse, dnorm(ftse, mean = mean(ftse), sd = sd(ftse)), col = "red")

> yvals <- dt((ftse - mu)/sigma, df = nu)/sigma
> lines(ftse, yvals, col = "blue")
```



Let’s practice!

Characteristics of volatile return series

Log-returns compared with iid data

- Can financial returns be modeled as **independent and identically distributed** (iid)?
- Random walk model for log asset prices
- Implies that future price behavior cannot be predicted
- Instructive to compare real returns with iid data
- Real returns often show **volatility clustering**

Let's practice!

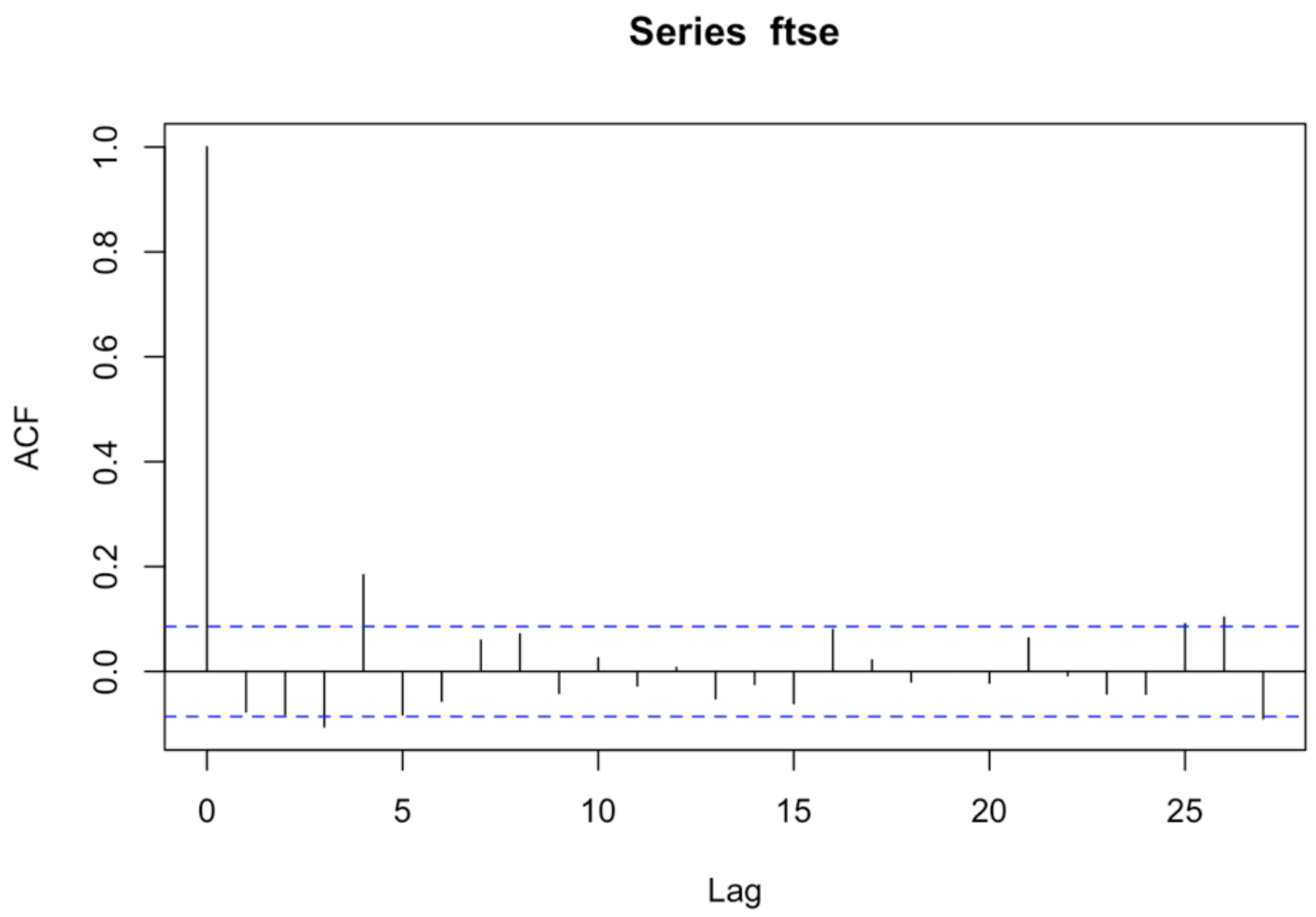
Estimating serial correlation

Sample autocorrelations

- **Sample autocorrelation function** (acf) measures correlation between variables separated by **lag** (k)
- **Stationarity** is implicitly assumed:
  - Expected return constant over time
  - Variance of return distribution always the same
  - Correlation between returns k apart always the same
- Notation for sample autocorrelation: $\hat{\rho}(k)$

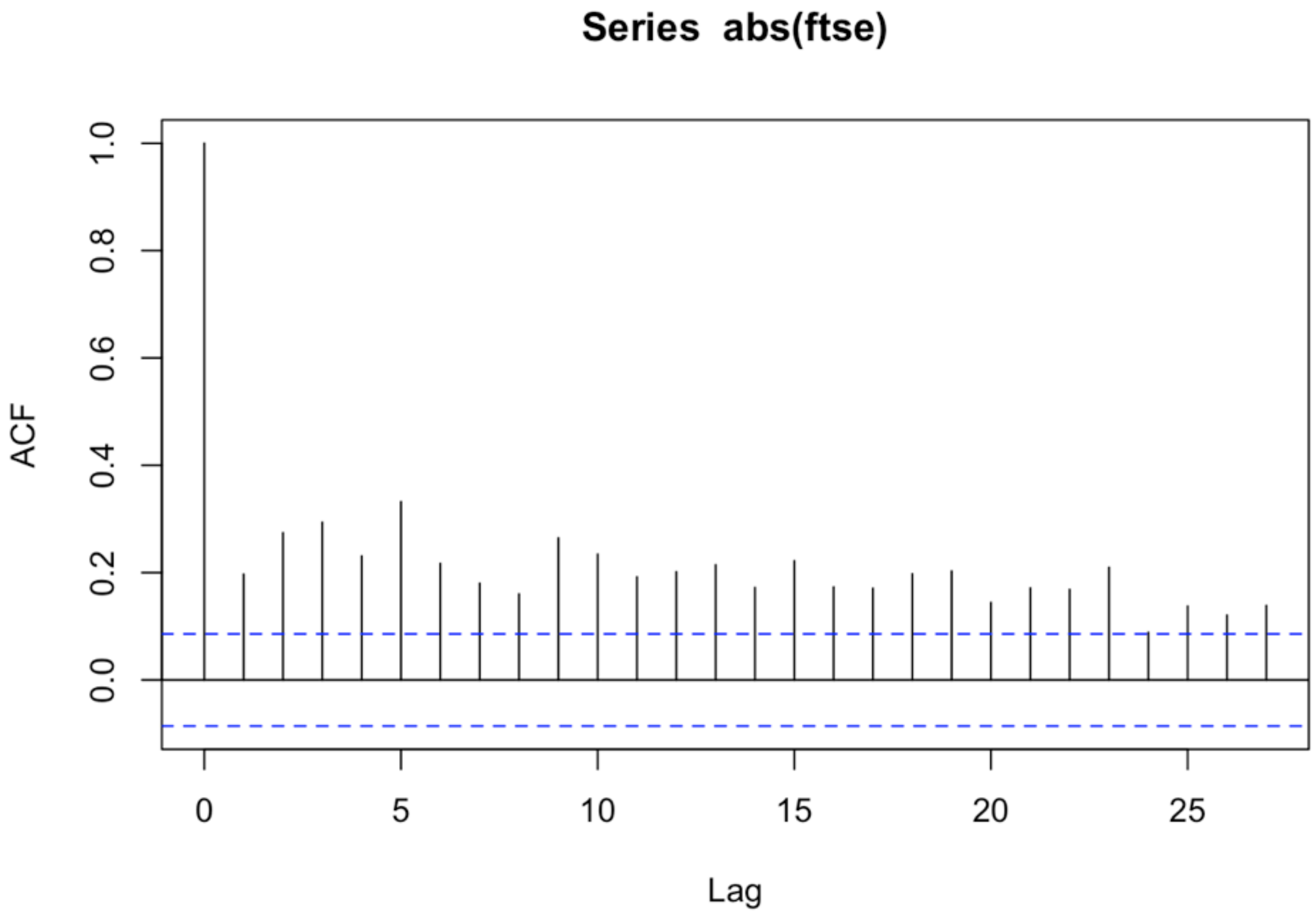
The sample acf plot or correlogram

```
> acf(ftse)
```



The sample acf plot or correlogram

```
> acf(abs(ftse))
```



Let's practice!

The Ljung-Box test

Testing the iid hypothesis with the Ljung-Box test

- Numerical test calculated from squared sample autocorrelations up to certain lag
- Compared with chi-squared distribution with k **degrees of freedom** (df)
- Should also be carried out on absolute returns

$$X^2 = n(n+2)\sum_{j=1}^{k}\frac{\hat{\rho}(j)^2}{n-j}$$

Example of Ljung-Box test

```
> Box.test(ftse, lag = 10, type = "Ljung")

	Box-Ljung test

data:  ftse
X-squared = 41.602, df = 10, p-value = 8.827e-06

> Box.test(abs(ftse), lag = 10, type = "Ljung")

	Box-Ljung test

data:  abs(ftse)
X-squared = 314.62, df = 10, p-value < 2.2e-16
```

Applying Ljung-Box to longer-interval returns

```
> ftse_w <- apply.weekly(ftse, FUN = sum)
> head(ftse_w, n = 3)
                  ^FTSE
2008-01-04 -0.01693075
2008-01-11 -0.02334674
2008-01-18 -0.04963134
```

```
> Box.test(ftse_w, lag = 10, type = "Ljung")

	Box-Ljung test

data:  ftse_w
X-squared = 18.11, df = 10, p-value = 0.05314

> Box.test(abs(ftse_w), lag = 10, type = "Ljung")

	Box-Ljung test

data:  abs(ftse_w)
X-squared = 34.307, df = 10, p-value = 0.0001638
```

Let's practice!

Looking at the extreme in financial time series

Extracting the extreme of return series

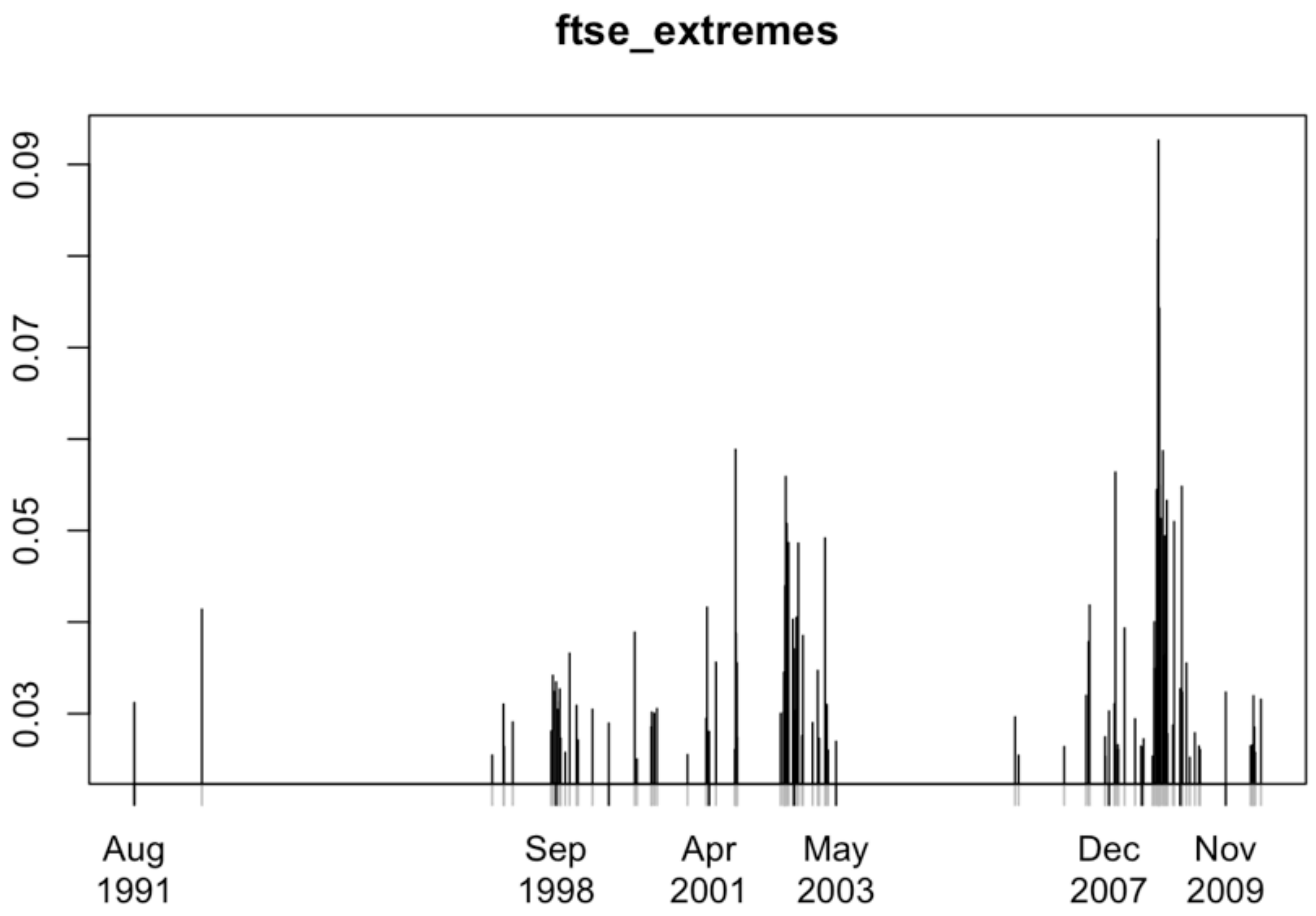
- Extract the most extreme negative log-returns exceeding 0.025

```
> ftse <- diff(log(FTSE))["1991-01-02/2010-12-31"]
> ftse_losses <- -ftse
> ftse_extremes <- ftse_losses[ftse_losses > 0.025]

> head(ftse_extremes)
                ^FTSE
1991-08-19 0.03119501
1992-10-05 0.04139899
1997-08-15 0.02546526
1997-10-23 0.03102717
> length(ftse_extremes)
[1] 115
```

- There are none from 1993-1996!

Plotting the extremes values

```
> plot(ftse_extremes, type = "h", auto.grid = FALSE)
```



Let's practice!

The stylized facts of return series

The stylized facts

1. Return series are heavier-tailed than normal, or **leptokurtic**
2. The volatility of return series appears to vary over time
3. Return series show relatively little serial correlation
4. Series of absolute returns show profound serial correlation
5. Extreme returns appear in clusters
6. Returns aggregated over longer periods tend to become more normal and less serially dependent

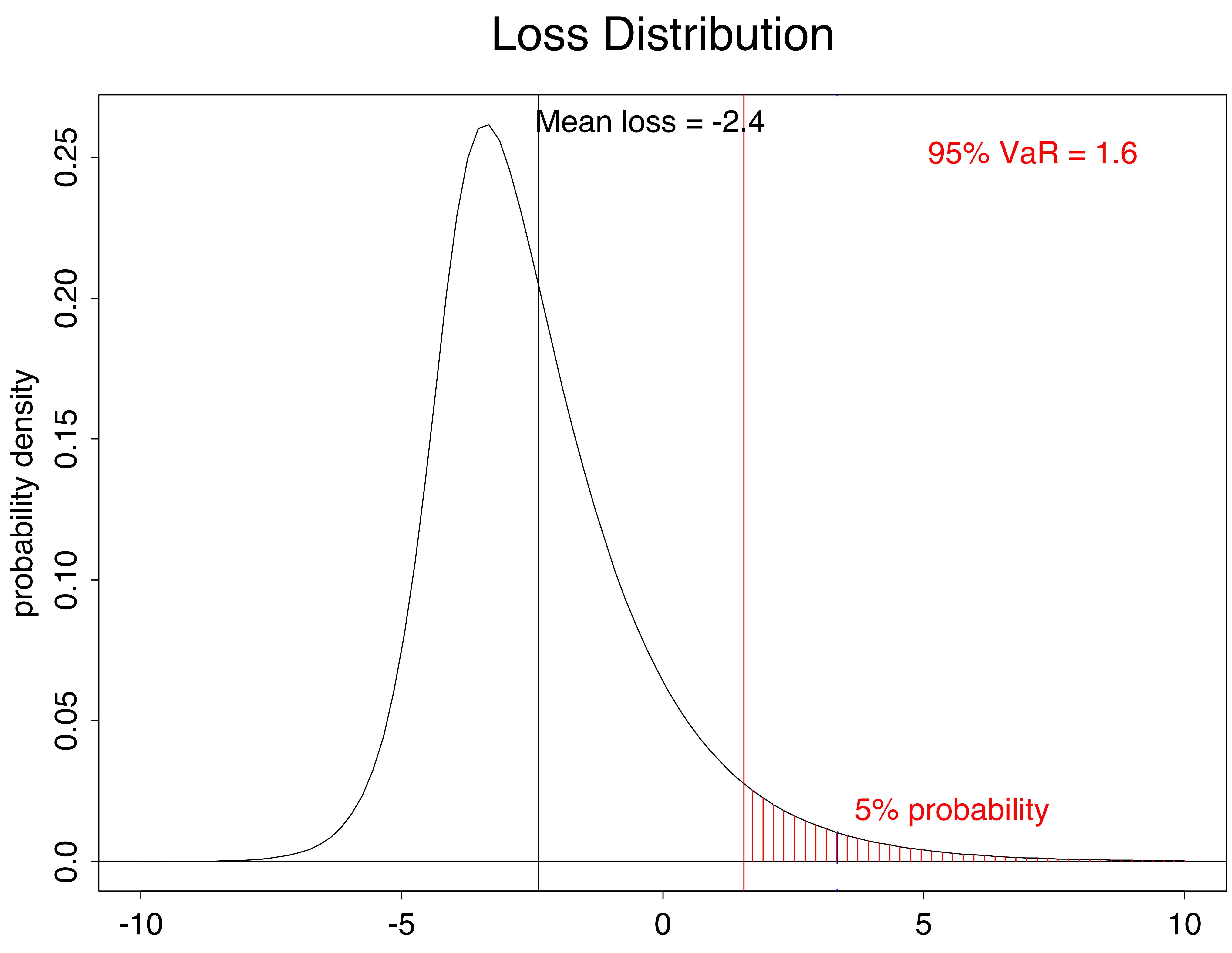
Let’s practice!

Value-at-risk and expected shortfall

Value-at-risk (VaR)

- Consider the distribution of losses over a fixed time period (day, week, etc.)
- α-VaR is the α-**quantile** of the loss distribution
- α known as **confidence level** (e.g. 95%, 99%)
- Should lose no more than α-VaR with probability α

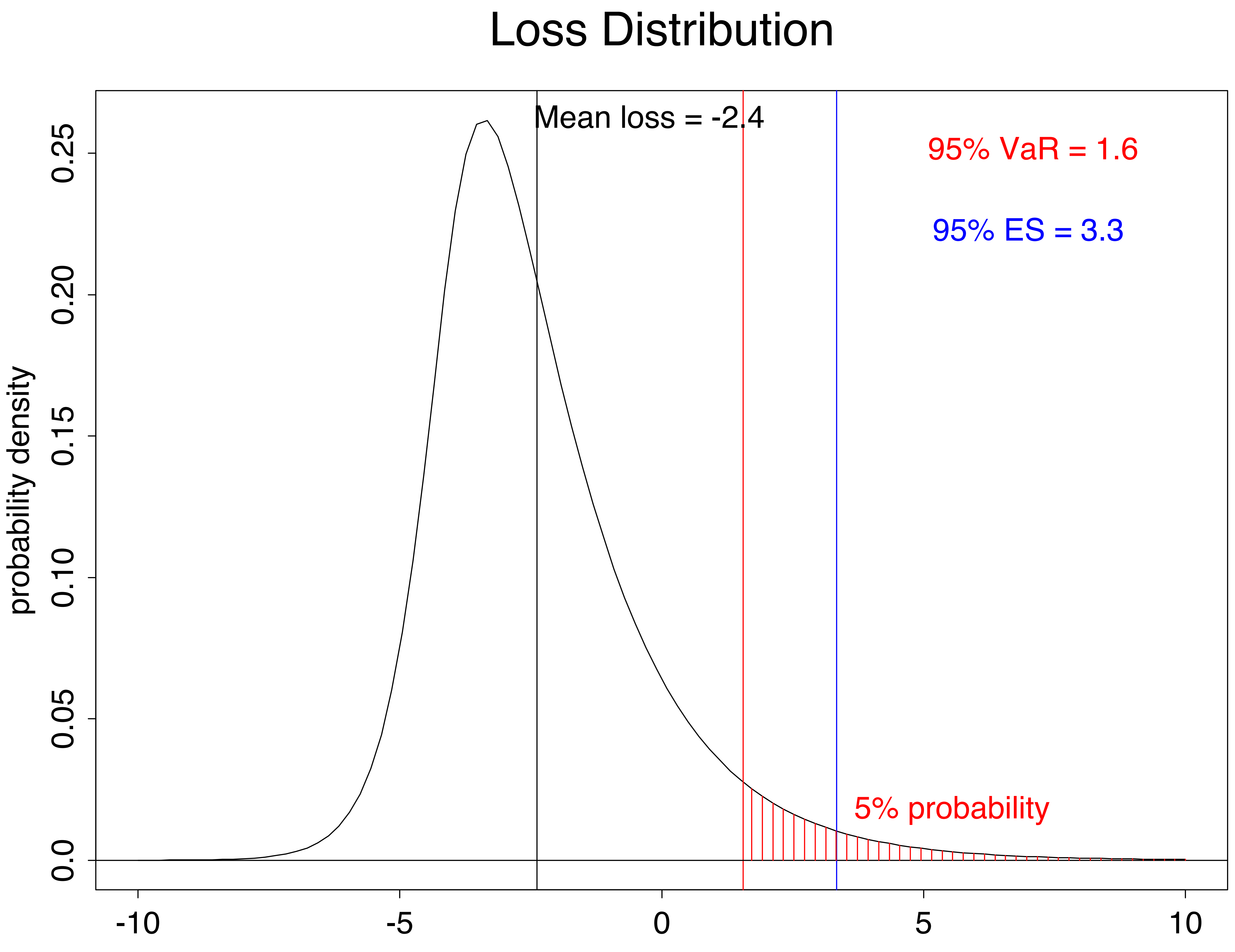
95% VaR illustrated



Expected shortfall (ES)

- Increasingly important in banking regulation
- Tail VaR (TVaR), conditional VaR (CVaR) or **expected shortfall** (ES)
- α-ES is expected loss given that loss exceeds α-VaR
- Expectation of tail of distribution

95% ES illustrated



Let's practice!

International equity portfolio example

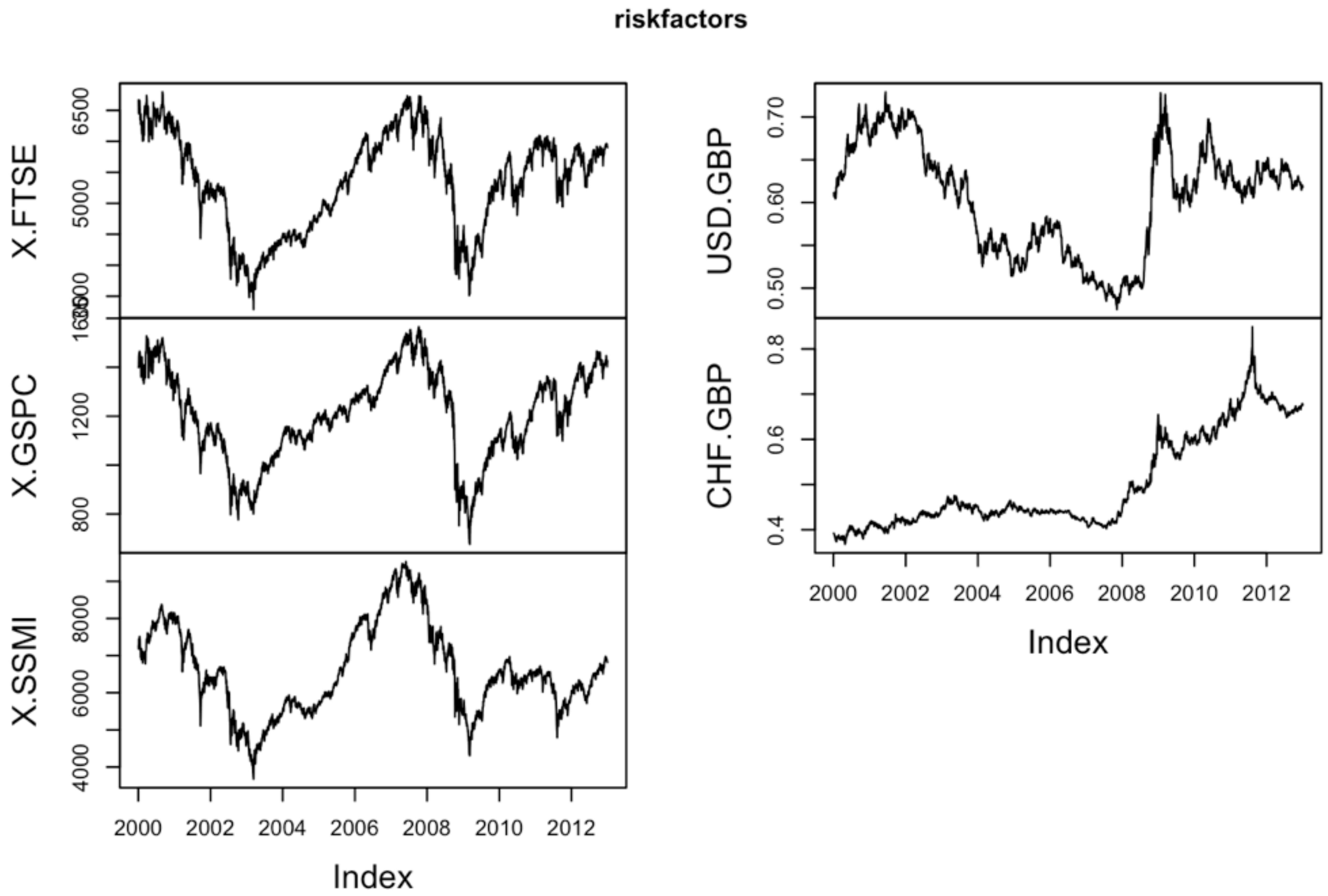
International equity portfolio

- Imagine a UK investor who has invested her wealth:
    - 30% FTSE, 40% S&P 500, 30% SMI
- 5 risk factors: FTSE, S&P 500 and SMI indexes, GBP/USD and GBP/CHF exchange rate

```
> riskfactors <- merge(FTSE, SP500, SMI, USD_GBP, CHF_GBP, all = FALSE)
["/2012-12-31", ]
```

Displaying the risk factors

```
> plot.zoo(riskfactors)
```



Historical simulation

- Simple method that is widely used in financial industry
- Resample historical risk-factor returns and examine their effect on current portfolio
- **Loss operator** shows effect of different risk-factor returns on the portfolio
- Loss operator functions will be provided in the exercises

Empirical estimates of VaR and ES

```
> losses <- rnorm(100)
> losses_o <- sort(losses, decreasing = TRUE)
> head(losses_o, n = 8)
[1] 1.836163 1.775163 1.745427 1.614479 1.602120 1.590034 1.483691
1.408354
> quantile(losses, 0.95)
     95%
1.590638
> qnorm(0.95)
[1] 1.644854
```

```
> mean(losses[losses > quantile(losses, 0.95)])
[1]  1.714671
> ESnorm(0.95)
[1] 2.062713
```

Let’s practice!

Option portfolio and Black-Scholes

European options and Black-Scholes

- European **call** option: gives right but not obligation to buy stock for price K at time T
- European **put** option: gives right but not obligation to sell stock for price K at T
- Value at time `t` < T depends on:
  - Stock price S, time to maturity T-`t`, interest rate `r`, annualized volatility σ or `sigma`
- Pricing by **Black-Scholes** formula

Pricing a first call option

```
> K <- 50
> T <- 2
> t <- 0
> S <- 40
> r <- 0.005
> sigma <- 0.25
> Black_Scholes(t, S, r, sigma, K, T, "call")
[1] 2.619183
> Black_Scholes(t, S, r, sigma*1.2, K, T, "call")
[1] 3.677901
```

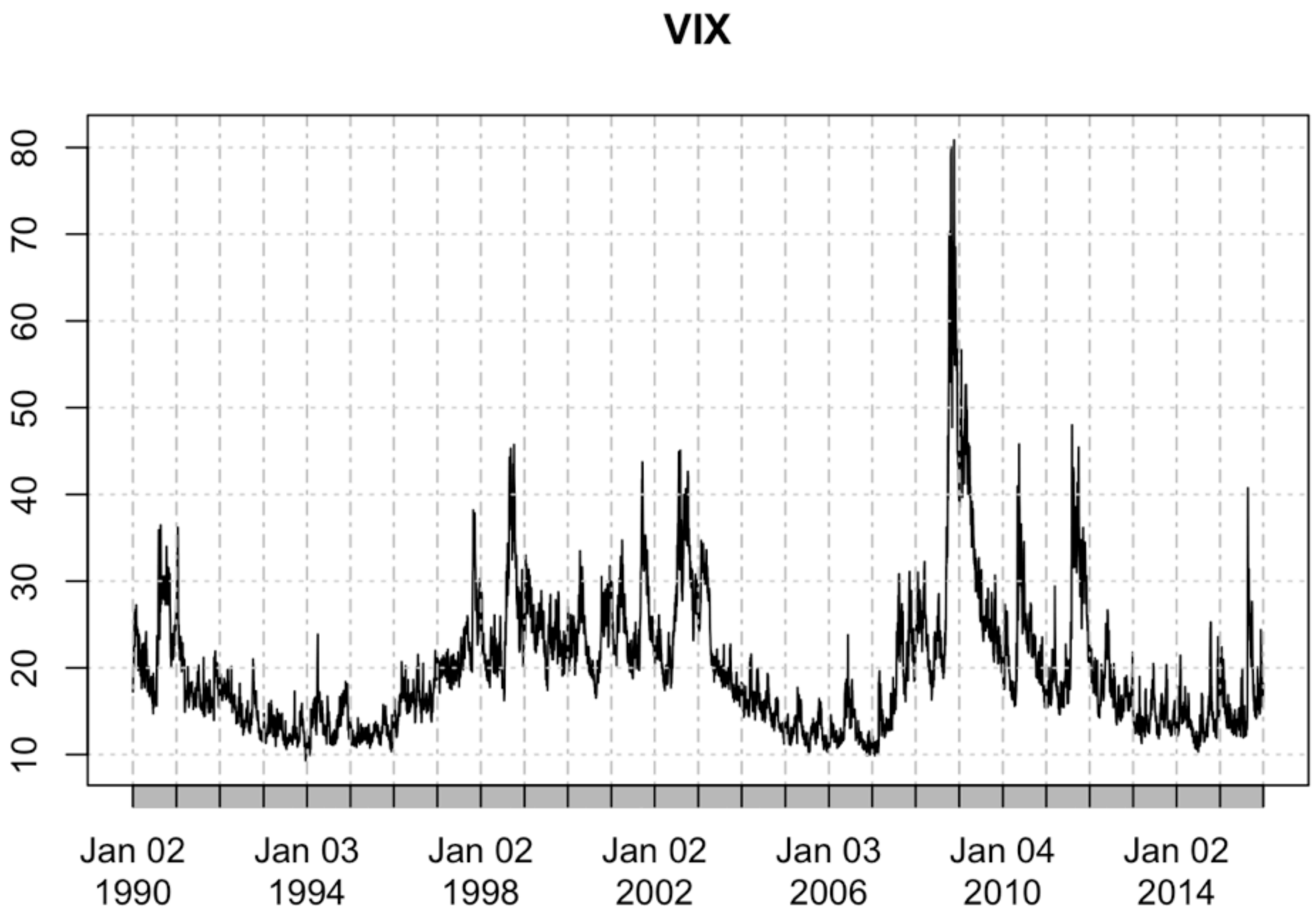
- Price increases with **volatility**
- Option above is **in-the-money**

Implied volatility X needs change

- Volatility not directly observable
- Market participants use **implied volatility**, the value of volatility implied by quoted option price

The VIX index

```
> plot(VIX)
```



Let's practice!

Historical simulation for the option example

Historical simulation

- Portfolio: single European call option on equity index
- Consider losses and profits over one day
- Changes to index value S, implied volatility σ and interest rate r affect value of portfolio
- We consider S and σ (and assume r stays constant)
- Create loss operator taking S and σ as input and giving the loss or profit as output

Estimating VaR and ES

- Apply loss operator `lossop()` to historical log-returns of S&P 500 and VIX to get simulated losses
- Estimate VaR by sample quantile as before
- Estimate ES by average of losses exceeding VaR estimate

Let's practice!

Wrap up

Not the end of the story...

Consider two things:

1. Can we improve risk sensitivity of VaR and ES estimates?
   - Filtered historical simulation, GARCH models, EWMA volatility filters
2. Can we improve simple empirical estimates of VaR and ES?
   - Parametric tail models, heavy-tailed distributions, extreme value theory

Thanks for taking the course!