Financial Risk Forecasting
Seminar Week 3: Distributions and Statistical Analysis

Jon Danielsson
London School of Economics

Version 4.0

3 Distributions and analysis

3.1 Why distributions and statistical analysis matter for financial risk

Financial returns follow non-normal distributions with fat tails and volatility clustering. Understanding these statistical properties is essential for accurate risk measurement - assuming normal distributions can severely underestimate tail risks. Statistical tests help validate model assumptions and detect structural breaks that affect risk forecasts.

These distribution concepts form the foundation for VaR calculations, stress testing and model validation techniques used throughout risk management.

For more detail, see Descriptive statistics in the R notebook.

3.2 The plan for this week

1. Work with distributions
2. Visualise, analyse and comment on the prices of a stock
3. Perform graphical analyses and statistical tests

3.3 Loading data and libraries

You may need to install some of those.

```
# Install packages if needed
# install.packages(c("tseries", "car", "lubridate", "moments"), repos = "https://cloud.r-p

library(tseries)
library(car)
library(lubridate)
library(moments)
```

```
load('Returns.RData')
load('Prices.RData')
load('NormalisedPrices.RData')
```

3.4 Links from the R notebook

Descriptive statistics

3.5 Work with distributions

We work extensively with statistical distributions, such as the normal, log-normal, Student-t, binomial, Bernoulli and Chi-square. R can, of course, handle many more distributions, but these are the ones we mostly use here. Each comes in 4 versions:

1. Density (pdf) — `d`
2. Distribution or cumulative density (cdf) — `p`
3. Quantiles — `q`
4. Random numbers — `r`

The function name is created by pre-fixing one of these four letters to the distribution name. For the normal, `dnorm, pnorm, qnorm, rnorm` and the t `dt, pt, qt, rt`.

We deal with random numbers in a later Seminar.

Here is an example showing how to plot the distributions over their domain.
```
x=seq(-3,3,length=1000)
z=seq(0,1,length=1000)
par(mfrow=c(2,2))
plot(x,dnorm(x), main="Normal Density",type='l')
plot(x,pnorm(x), main="Cumulative Density",type='l')
plot(z,qnorm(z), main="Normal Quantile",type='l')
```

Normal Density

Cumulative Density

Normal Quantile
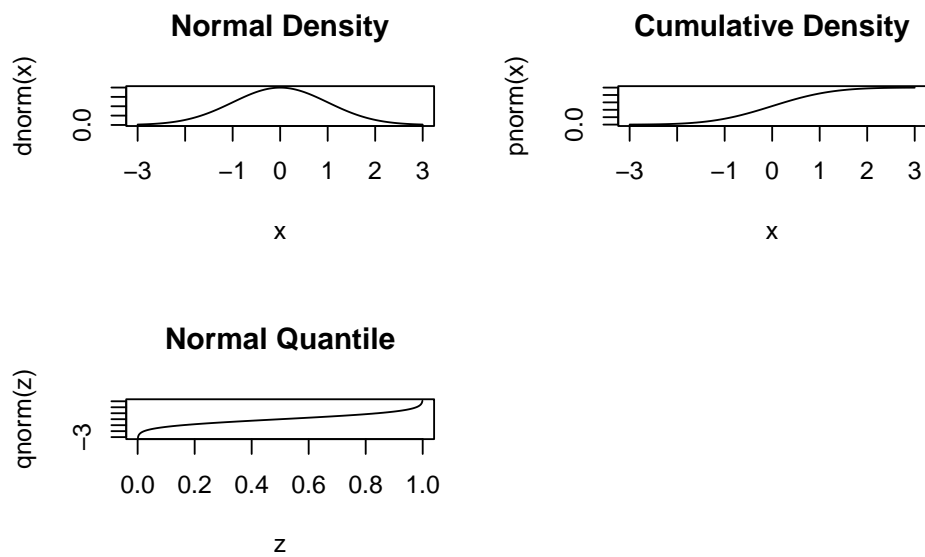
3.6 Comparing the normal distribution with the Student-t

The Student-t distribution has fatter tails than the normal.

```r
x = seq(-3, 3, length=1000)
normal = dnorm(x)
st2 = dt(x, df = 2)
st3 = dt(x, df = 3)
st10 = dt(x, df = 10)
plot(x, normal, type = "l", main = "Comparing distributions", col = 1, xlab = "x", ylab =
lines(x, st2, col = 2)
lines(x, st3, col = 3)
lines(x, st10, col = 4)
legend("topright",
    legend = c("Normal", "T - 2 df", "T - 3 df", "T - 10 df"),
    col = c(1:4),
    lty=1,
    bty='n'
)
```

Comparing distributions
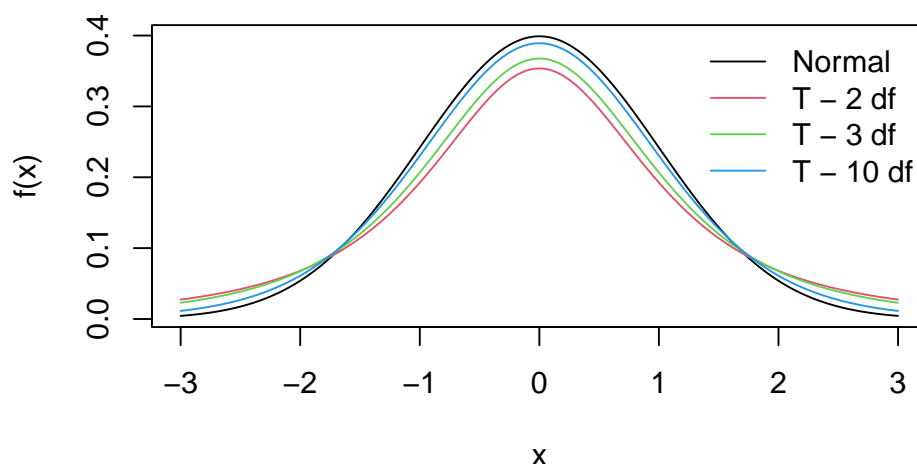
These distributions are important in financial analysis because:

- **Normal distribution** is the traditional assumption in many financial models and risk calculations
- **Student-t distribution** better captures the "fat tails" often seen in financial returns — extreme losses and gains happen more frequently than the normal distribution predicts
- The choice between these distributions affects risk calculations, portfolio optimisation and derivative pricing

3.7 Applying distribution concepts to real financial data

Now that we understand the theoretical distributions commonly used in finance, we can apply these concepts to analyse real stock price data. We will examine how actual returns compare to theoretical distributions and identify periods where normal distribution assumptions break down.

3.8 Visualising and commenting on the price of a stock

```
head(Prices)
```

```
            date    GSPC     IXIC    AAPL    MSFT     JPM         C      XOM
18078 2000-01-03 1455.22  4131.15  0.8410 35.6930 23.1248 209.0070 18.1273
18079 2000-01-04 1399.42  3901.69  0.7702 34.4873 22.6175 196.1906 17.7801
18080 2000-01-05 1402.11  3877.54  0.7814 34.8509 22.4778 204.0776 18.7494
18081 2000-01-06 1403.45  3727.13  0.7138 33.6834 22.7970 213.9364 19.7187
18082 2000-01-07 1441.47  3882.62  0.7476 34.1237 23.2158 212.9506 19.6608
18083 2000-01-10 1457.60  4049.67  0.7345 34.3724 22.8169 212.2112 19.3859
          MCD       GE   NVDA
18078 21.2780 130.0416 0.0894
18079 20.8417 124.8399 0.0870
18080 21.1774 124.6232 0.0842
18081 20.8753 126.2894 0.0787
```

```
18082 21.4123 131.1795 0.0800
18083 21.5130 131.1253 0.0826
```

```
head(NormalisedPrices)
```

```
            date      GSPC      IXIC      AAPL      MSFT       JPM         C
18078 2000-01-03 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
18079 2000-01-04 0.9616553 0.9444561 0.9158145 0.9662203 0.9780625 0.9386796
18080 2000-01-05 0.9635038 0.9386103 0.9291320 0.9764071 0.9720214 0.9764151
18081 2000-01-06 0.9644246 0.9022016 0.8487515 0.9436976 0.9858247 1.0235849
18082 2000-01-07 0.9905513 0.9398400 0.8889417 0.9560334 1.0039352 1.0188683
18083 2000-01-10 1.0016355 0.9802767 0.8733650 0.9630011 0.9866853 1.0153306
            XOM       MCD        GE      NVDA
18078 1.0000000 1.0000000 1.0000000 1.0000000
18079 0.9808466 0.9794953 0.9599997 0.9731544
18080 1.0343184 0.9952721 0.9583333 0.9418345
18081 1.0877902 0.9810743 0.9711462 0.8803132
18082 1.0845962 1.0063117 1.0087503 0.8948546
18083 1.0694312 1.0110443 1.0083335 0.9239374
```

3.8.1 GE case

Jack Welch, who was the CEO of GE for 20 years, retired on September 7, 2001. He was considered to be one of the most valuable CEOs of all time. Add a vertical line on our plot to reflect this:

Find what the maximum price of GE was during Jack Welch's tenure and when it was reached:

```
plot(Prices$date,Prices$GE, type = "l", main = "Price of GE")
# Limit to first 1000 observations to focus on the period when Welch was leaving
maxPrice=max(Prices$GE[1:1000])
MaxDate=Prices$date[Prices$GE == maxPrice]
abline(v = ymd(20010907), lwd = 2, col = "red")
abline(v = MaxDate, lwd = 2, col = "blue")
text(MaxDate, maxPrice, pos=4,paste("Highest price\nuntil 2025", maxPrice))
```

Price of GE
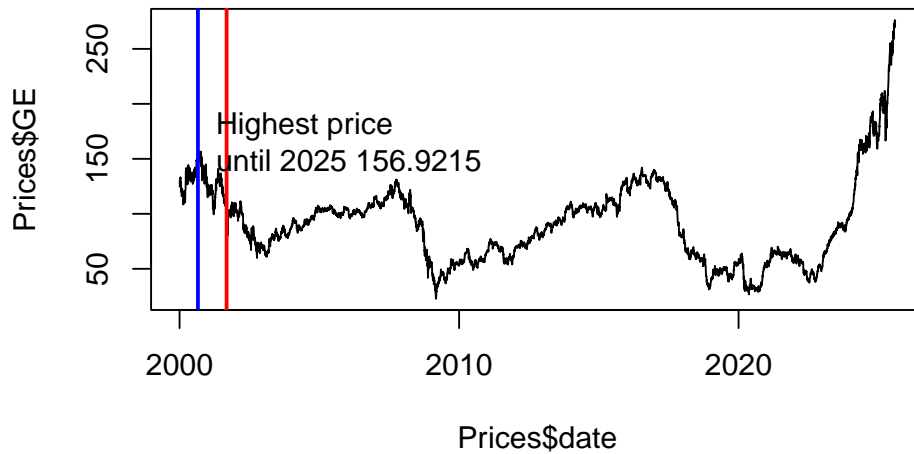
3.8.2 Zoom into the crisis in 2008.

To compare how different stocks performed during the crisis, we normalise all prices to start at 1. This shows relative performance regardless of absolute price levels — which stocks declined most and which recovered fastest.

```r
Prices.crisis= Prices[Prices$date>ymd(20080101) & Prices$date<ymd(20090701),]
for(i in 2:ncol(Prices.crisis)) {
  Prices.crisis[,i] = Prices.crisis[,i]/Prices.crisis[1,i]
  }
matplot(Prices.crisis$date, Prices.crisis[,2:ncol(Prices.crisis)], col=1:(ncol(Prices.cris
legend("bottomleft",
legend=names(Prices.crisis)[2:ncol(Prices.crisis)],
lty=1,
col=1:(ncol(Prices.crisis)-1),
bty='n',ncol=2
)
```
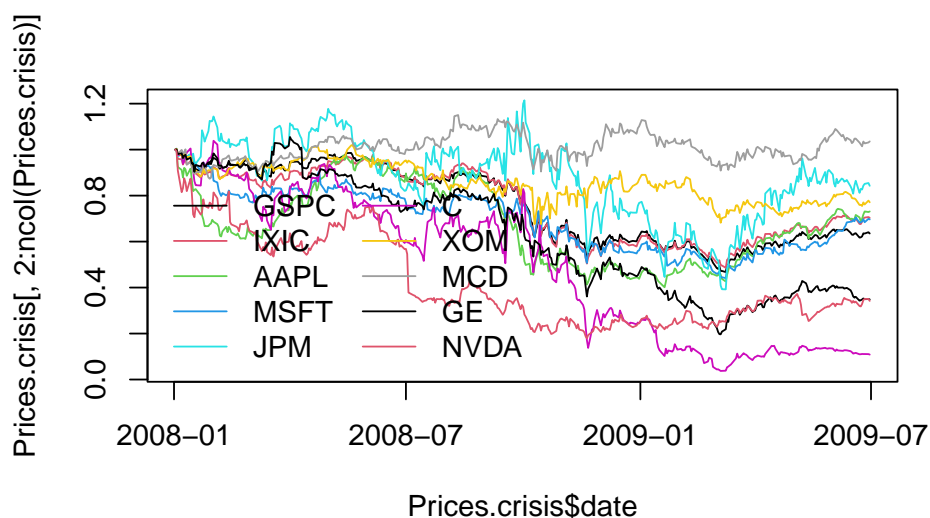
Prices.crisis$date

Now apply the same crisis period analysis to returns data to understand volatility patterns during the crisis:

```
Returns.crisis = Returns[Returns$date>ymd(20080101) & Returns$date<ymd(20090701),]
matplot(Returns.crisis$date, Returns.crisis[,2:ncol(Returns.crisis)],
        col=1:8, type='l', lty=1,
        main="Returns during 2008 Crisis",
        xlab="Date", ylab="Returns")
legend("bottomleft",
legend=names(Returns.crisis)[2:ncol(Returns.crisis)],
lty=1,
col=1:8,
bty='n',ncol=2
)
```
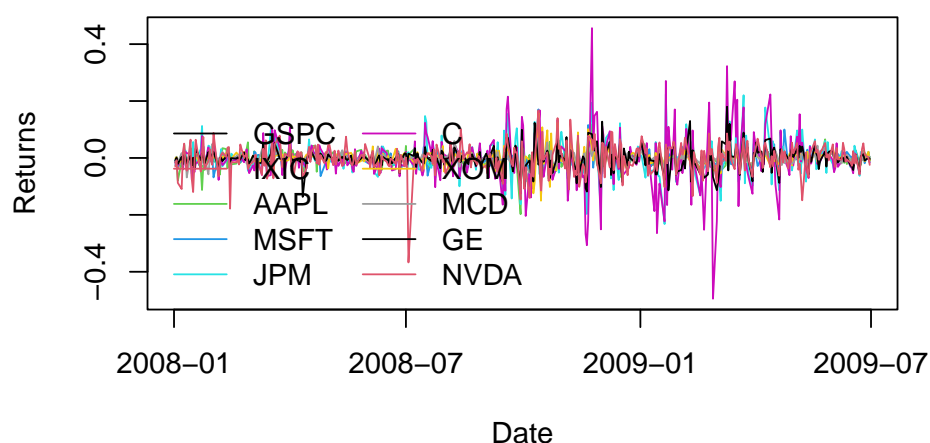

This shows the volatility clustering during the crisis — periods of high volatility tend to be followed by more high volatility.

3.8.3 Graphical analyses and statistical tests

We can do some of the basic statistical and graphical analysis shown at the start of the course. See the statistics chapter in the risk forecasting notebook.

Pick JP Morgan

```
y=Returns$JPM
```

First, print some summary statistics and run the Jarque Bera and Box tests.

```
mean(y)
```

```
[1] 0.0003821615
```

```
sd(y)
```

```
[1] 0.02330617
```

```
skewness(y)
```

```
[1] 0.2136689
```

```
kurtosis(y)
```

```
[1] 17.52491
```

```
jarque.bera.test(y)
```

```

	Jarque Bera Test

data:  y
X-squared = 56651, df = 2, p-value < 2.2e-16
```

```
Box.test(y, type = "Ljung-Box")
```

```

	Box-Ljung test

data:  y
X-squared = 43.907, df = 1, p-value = 3.443e-11
```

```
Box.test(y^2, type = "Ljung-Box")
```

```

	Box-Ljung test

data:  y^2
X-squared = 806.15, df = 1, p-value < 2.2e-16
```
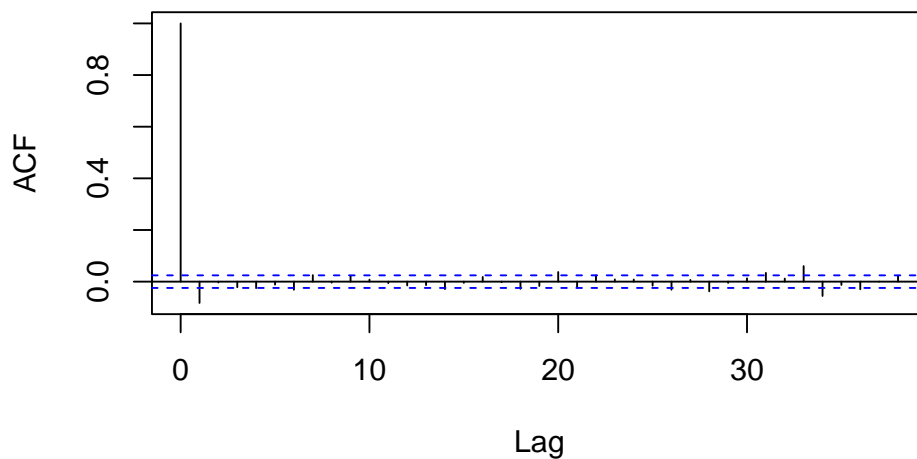
Then, plot the autocorrelation function of returns and returns squared. What information does the latter plot provide?
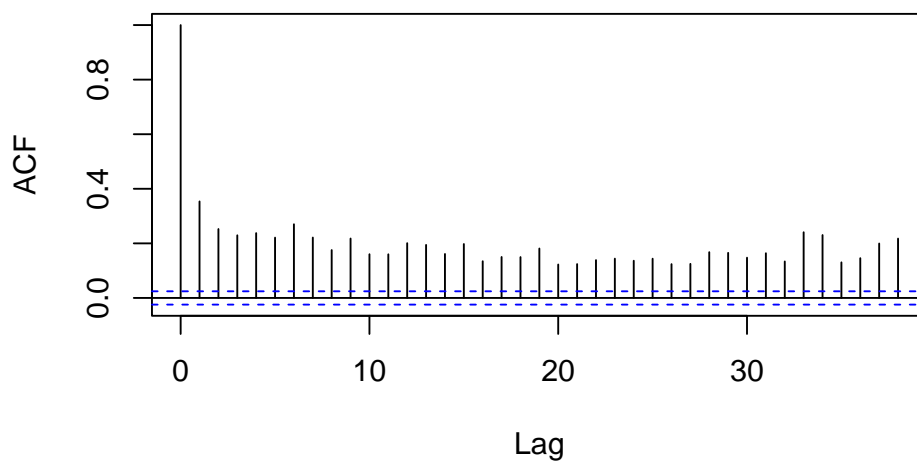
```
acf(y, main = "Autocorrelation of returns")
```

Autocorrelation of returns


```
acf(y^2, main = "Autocorrelation of returns squared")
```
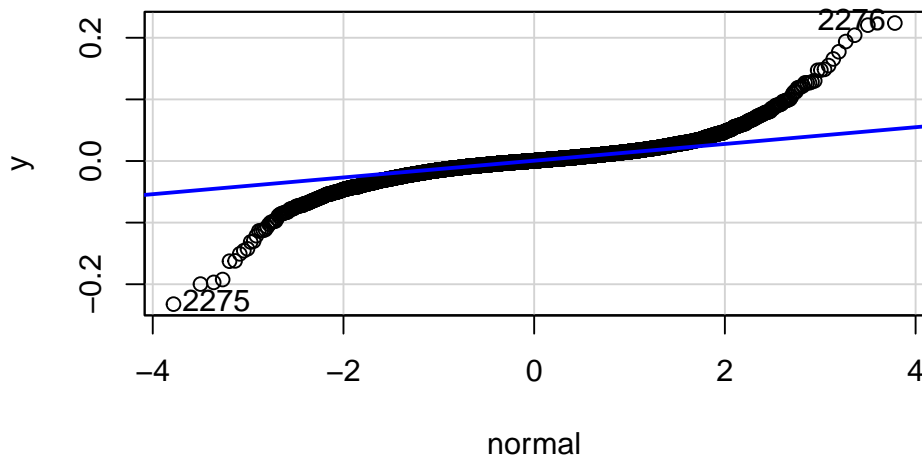
Autocorrelation of returns squared


These ACF plots look like they could be nicer. The notebook shows some better alternatives.

Finally, the QQ plot can be informative about the distribution of the returns, especially in the tails. Note the difference between the lower and upper tails.
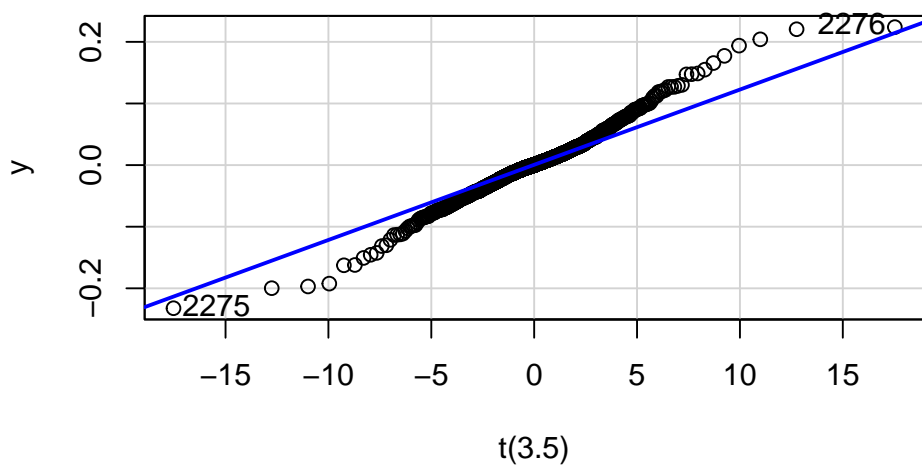
```
x=qqPlot(y, distribution = "norm", envelope = FALSE,xlab="normal")
```
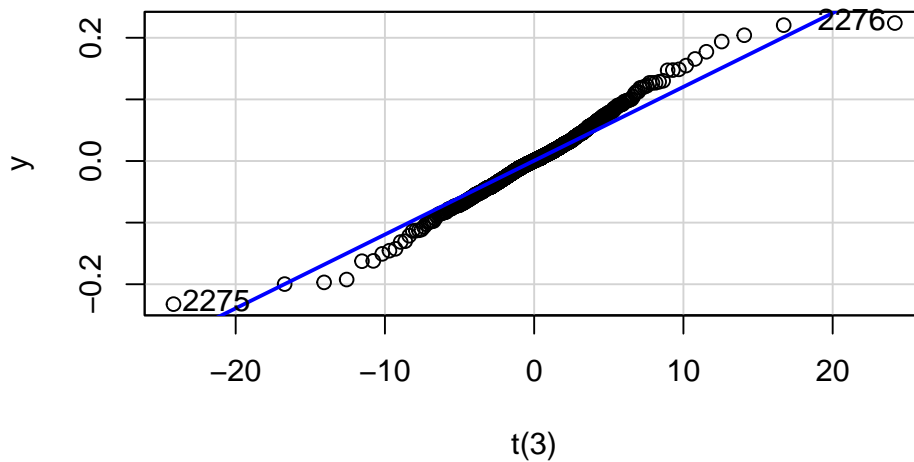
9

normal

```
x=qqPlot(y, distribution = "t", df = 4, envelope = FALSE,xlab="t(4)")
```


t(4)
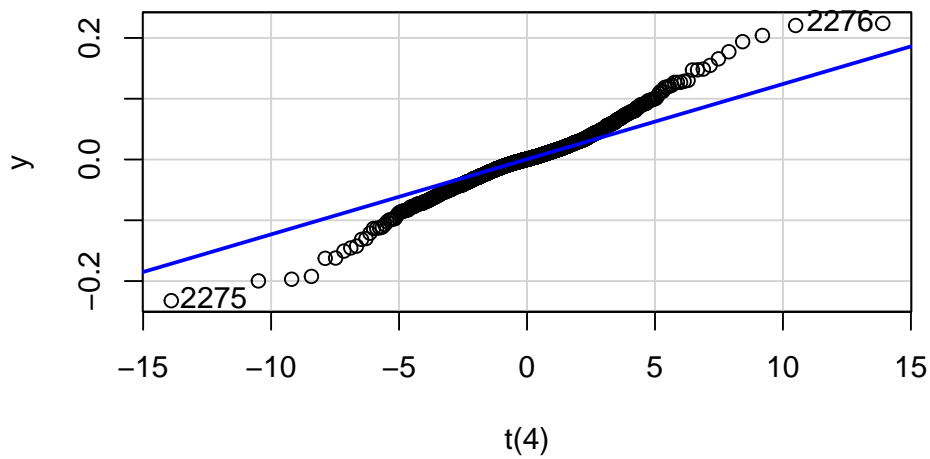
```
x=qqPlot(y, distribution = "t", df = 3.5, envelope = FALSE,xlab="t(3.5)")
```


t(3.5)

```
x=qqPlot(y, distribution = "t", df = 3, envelope = FALSE,xlab="t(3)")
```

3.9 Recap

3.9.1 In this seminar, we have covered:

- Working with statistical distributions in R:
  - Density functions (dnorm, dt)
  - Cumulative distribution functions (pnorm, pt)
  - Quantile functions (qnorm, qt)
  - Comparing normal and Student-t distributions
- Visualizing and analyzing stock price data:
  - Adding reference lines for important events
  - Finding and marking maximum/minimum values
  - Creating crisis period analysis
- Statistical testing and graphical analysis:
  - Calculating summary statistics (mean, sd, skewness, kurtosis)
  - Testing for normality and autocorrelation
  - Creating ACF plots for returns and squared returns
  - Using QQ plots to assess distributional assumptions

Some new functions used:

- `dnorm()` — normal density function
- `pnorm()` — normal cumulative distribution function
- `qnorm()` — normal quantile function
- `dt()` — Student-t density function
- `seq()` — generate sequences of numbers
- `par()` — set graphical parameters
- `text()` — add text annotations to plots
- `mean()` — calculate arithmetic mean
- `sd()` — calculate standard deviation
- `max()` — find maximum value
- `abline()` — add straight lines to plots (vertical, horizontal or sloped)
- `skewness()` — measure asymmetry of the distribution
- `kurtosis()` — measure tail heaviness of the distribution
- `jarque.bera.test()` — test for normality based on skewness and kurtosis
- `Box.test()` — test for autocorrelation in time series

- `acf()` — plot autocorrelation function
- `qqPlot()` — quantile-quantile plot for comparing distributions

3.10   Optional exercises

1. **Summary statistics comparison**:
   - Make a table with the key summary statistics for all stocks in our sample
   - Include mean, sd, skewness, kurtosis, min and max returns
   - Which stock has the most extreme skewness? What does this tell you?
2. **Visualization with statistics**:
   - Make a plot for each stock and put the key sample statistics on each figure
   - Use `text()` or `mtext()` to add the statistics to the plots
   - Create a 3x3 grid of plots using `par(mfrow)`
3. **Distribution analysis**:
   - For each stock, test if returns follow a normal distribution using Jarque-Bera test
   - Create a summary table showing which stocks reject normality at 5% level
   - Plot histograms with overlaid normal curves for comparison
4. **Tail behaviour investigation**:
   - Compare the tail behaviour of different stocks using QQ plots
   - Which stocks show the fattest tails?
   - Try fitting different degrees of freedom for the Student-t distribution
5. **Autocorrelation patterns**:
   - Create ACF plots for all stocks' squared returns in a single figure
   - Which stocks show the strongest volatility clustering?
   - Test for autocorrelation at lag 20 using Box-Ljung test
6. **Crisis period analysis**:
   - Extend the 2008 crisis analysis to include COVID-19 (March-April 2020)
   - Compare how different sectors performed in each crisis
   - Calculate maximum drawdowns during each period
7. **Rolling statistics**:
   - Calculate 60-day rolling skewness and kurtosis for one stock
   - Plot these over time - do they spike during crises?
   - Identify periods of abnormal distribution characteristics
8. **Sector comparison**:
   - Group stocks by sector (tech, finance, energy, etc.)
   - Compare average skewness and kurtosis by sector
   - Which sector has returns closest to normal distribution?