Faculty of Management

Financial Data Analysis (MQIM 6602)

Lectures: Mondays. 8:30-11:20am, on MS Teams

Winter 2021

Azam Shamsi Zamenjani

December 16, 2020

1 Introduction

A time series is a stochastic process $\{..., X_{t-1}, X_t, X_{t+1}, ...\}$ consisting of a random variable indexed by time: An ordered sequence of values of a variable at equally spaced time intervals (annual, monthly, weekly,daily, high-frequency data, ...). The observations in time series are almost never independent. Usually, one observation is correlated with the previous observation (e.g., cross-sectional unemployment and time series unemployment rate). This dependency in some sense makes some things easier: In time series, we watch something unfold slowly over time. If the economy changes slowly, then we can use the past as a useful guide to the future. Want to know what next month's unemployment rate will be? It will very likely be close to this month's rate. And it will be changing by roughly the same amount as it changed in the recent past.

Financial time series (FTS) analysis is concerned with theory and practice of asset valuation over time, for example stock prices, exchange rates,...

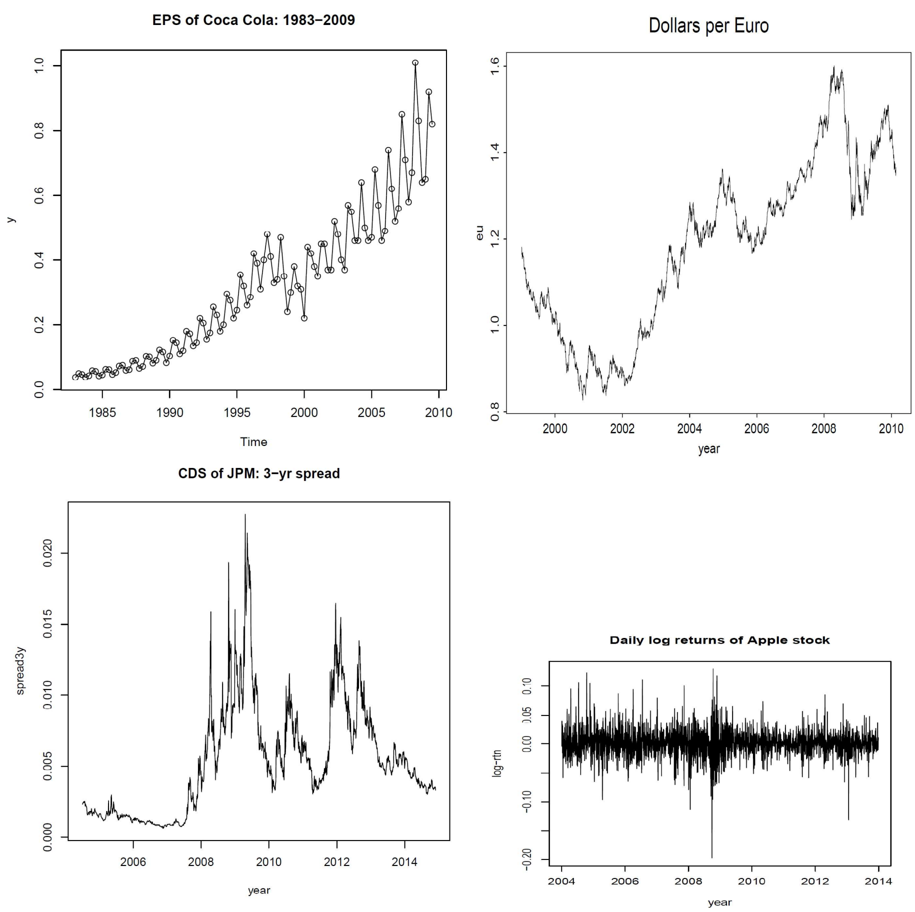

Figure 1: Sample Time Series.

Comparison with other time series analyses: Highly related, but with some added uncertainty, because FTS must deal with the ever-changing business & economic environment and the fact that volatility is not directly observed.

Objective of the course

- to learn ways to get financial information from web directly and to process the information.
- to provide some basic knowledge of financial time series data such as skewness, heavy tails, and measure of dependence between asset returns.
- to introduce some statistical tools & econometric models useful for analyzing these series.
- to gain experience in analyzing FTS.
- to study methods for assessing market risk, credit risk, and expected loss. The methods discussed include Value at Risk, expected shortfall, and tail dependence.
- to analyze high-dimensional asset returns, including co-movement
- to introduce Bayesian Econometrics and MCMC techniques

Examples of financial time series

1. Daily log returns of Apple stock: 2004 to 2013 (10 years)
2. The VIX index
3. CDS spreads: Daily 3-year CDS spreads of JP Morgan from July 20, 2004 to September 19, 2014
4. Quarterly earnings of Coca-Cola Company: 1983-2009
5. US monthly interest rates (3m & 6m Treasury bills)
6. Exchange rate between US Dollar vs Euro

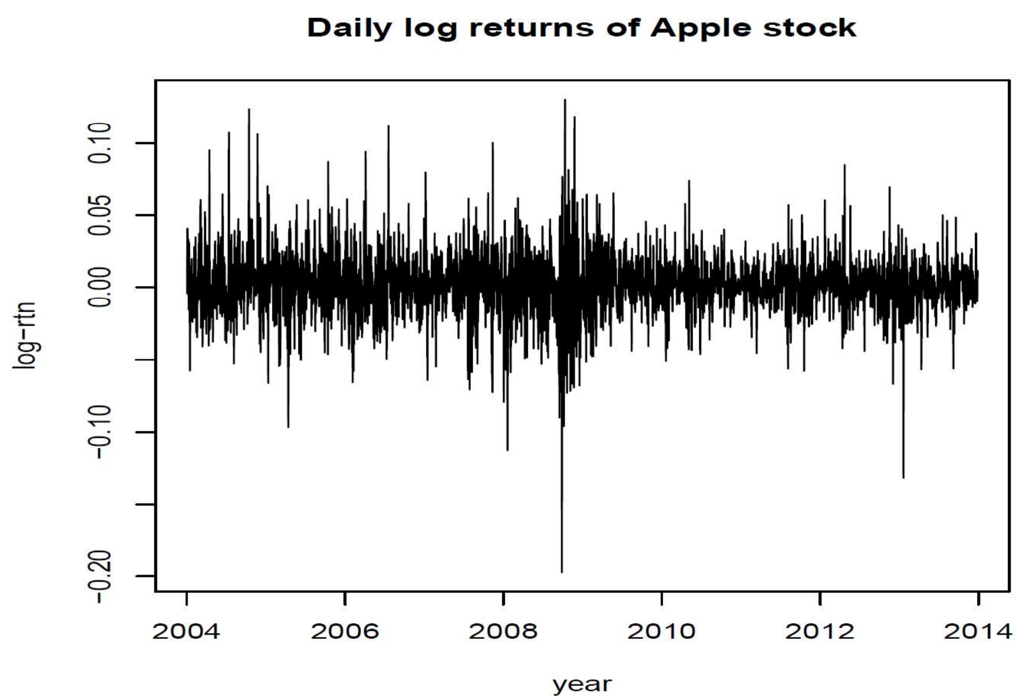

Figure 2: Daily log returns of Apple stock from 2004 to 2013.

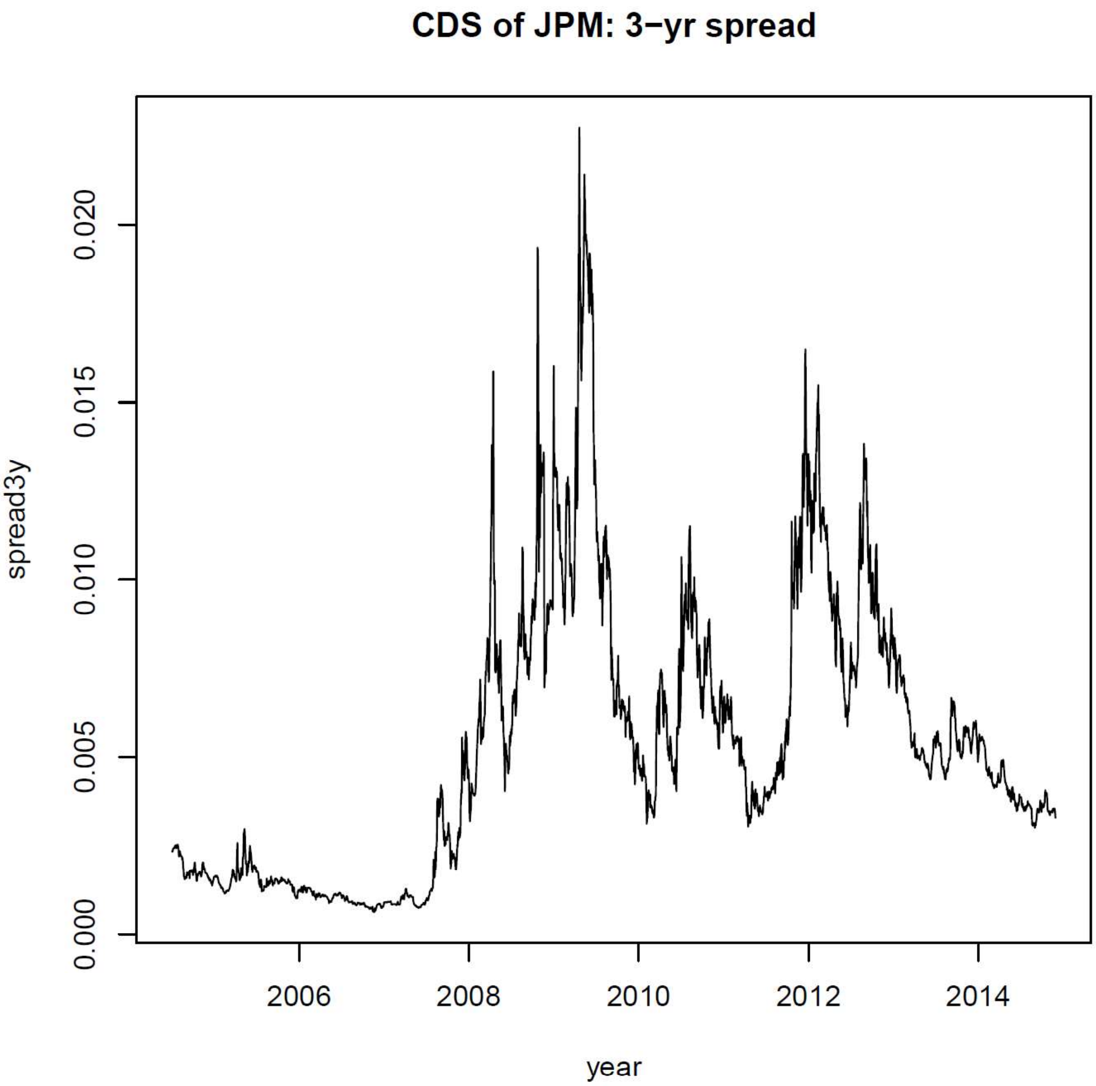

Figure 3: Time plot of daily 3-year CDS spreads of JPM: from July 20, 2004 to September 19, 2014.

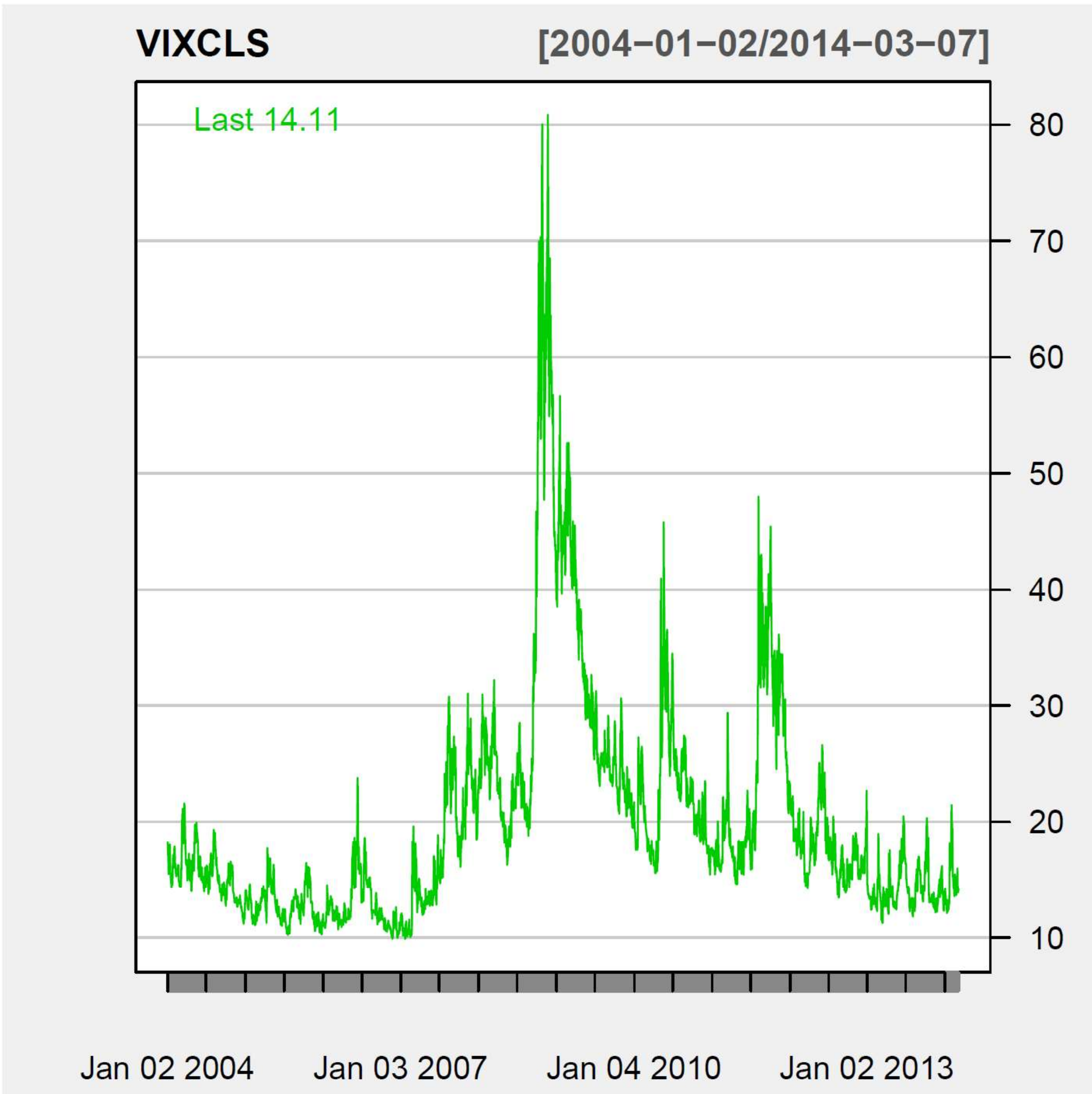

Figure 4: CBOE Vix index: January 2, 2004 to March 7, 2014.

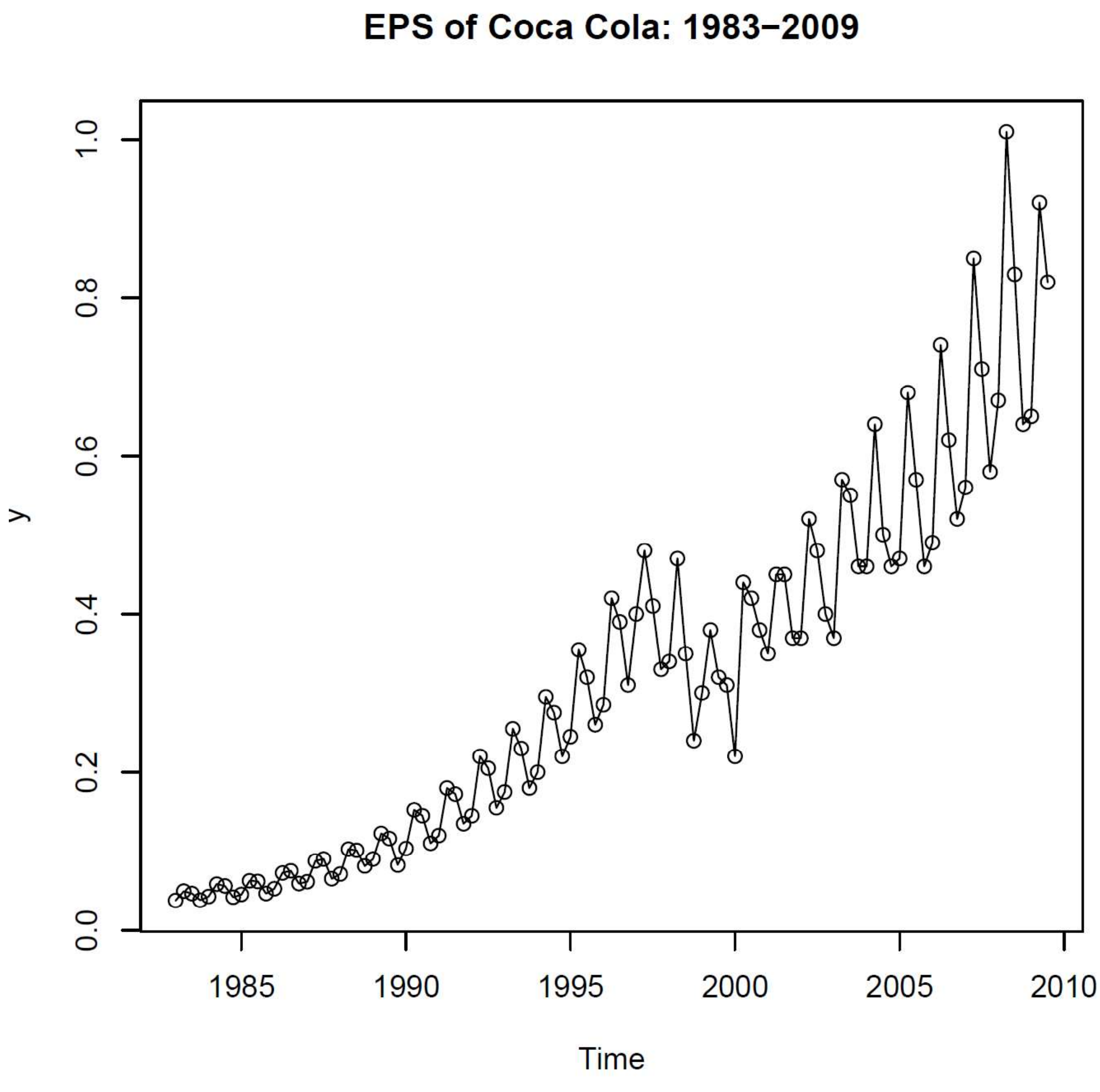

Figure 5: Quarterly earnings per share of Coca-Cola Company.

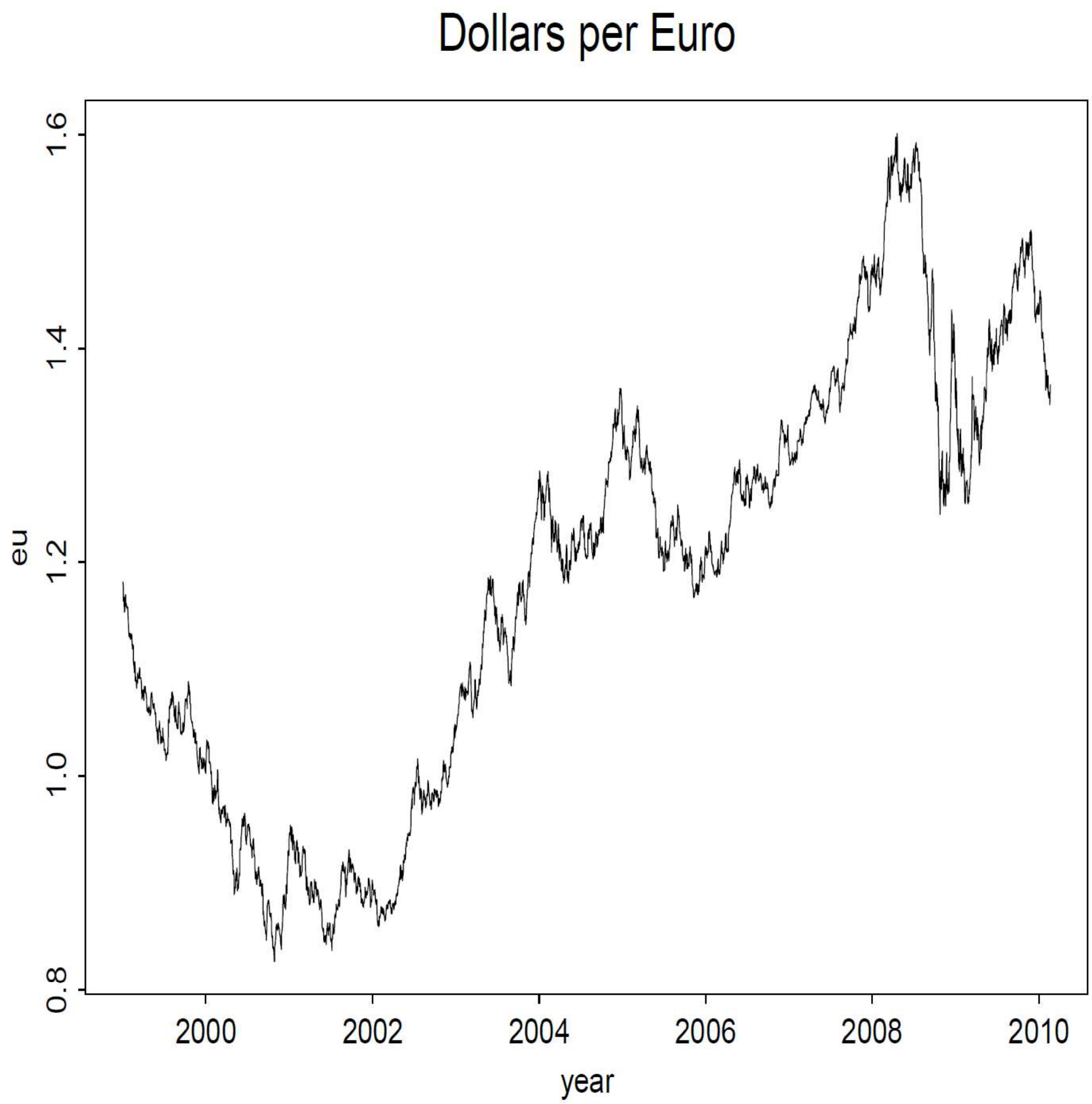

Figure 6: Daily Exchange Rate: Dollars per Euro.

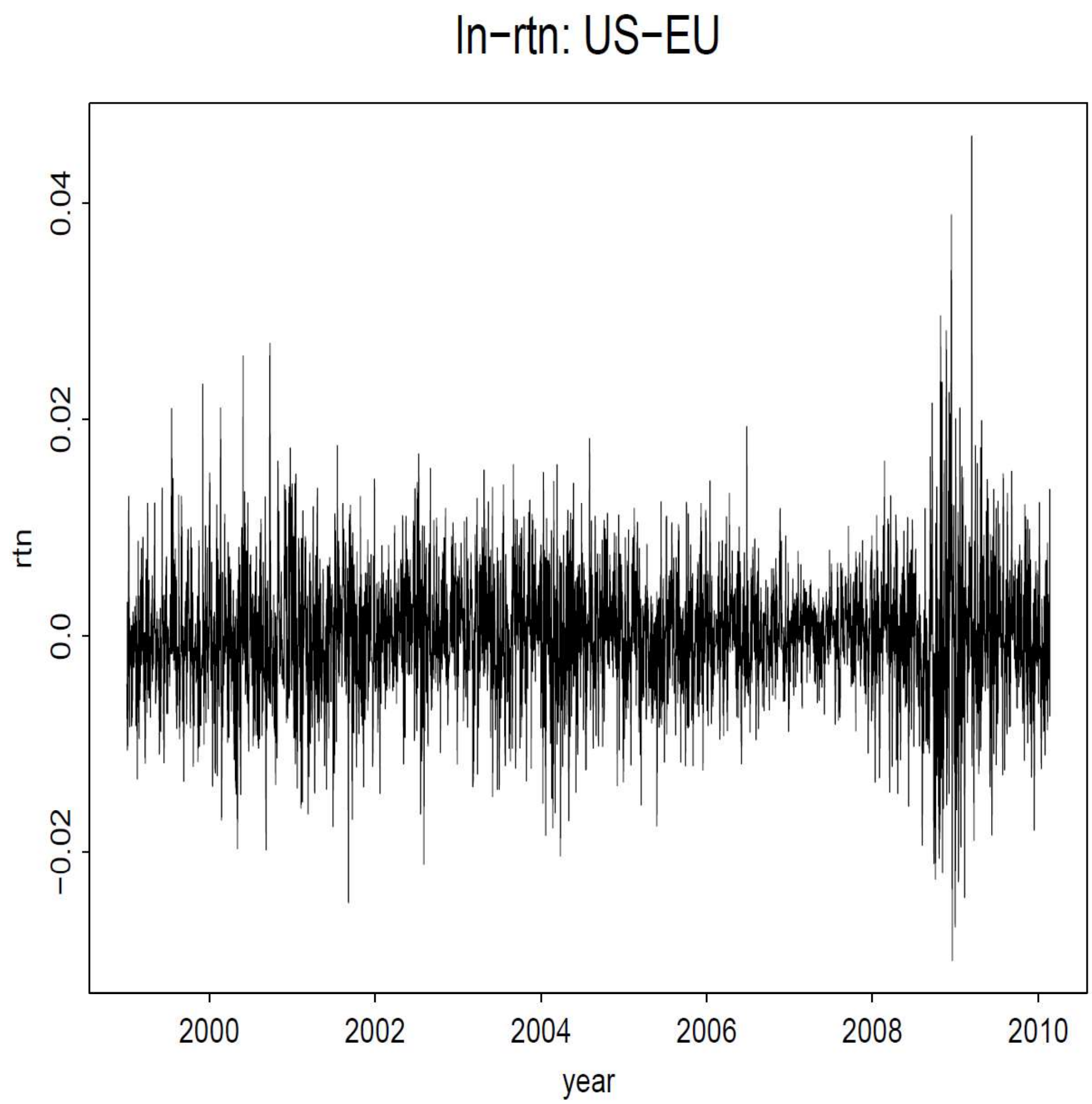

Figure 7: Daily log returns of FX (Dollar vs Euro).

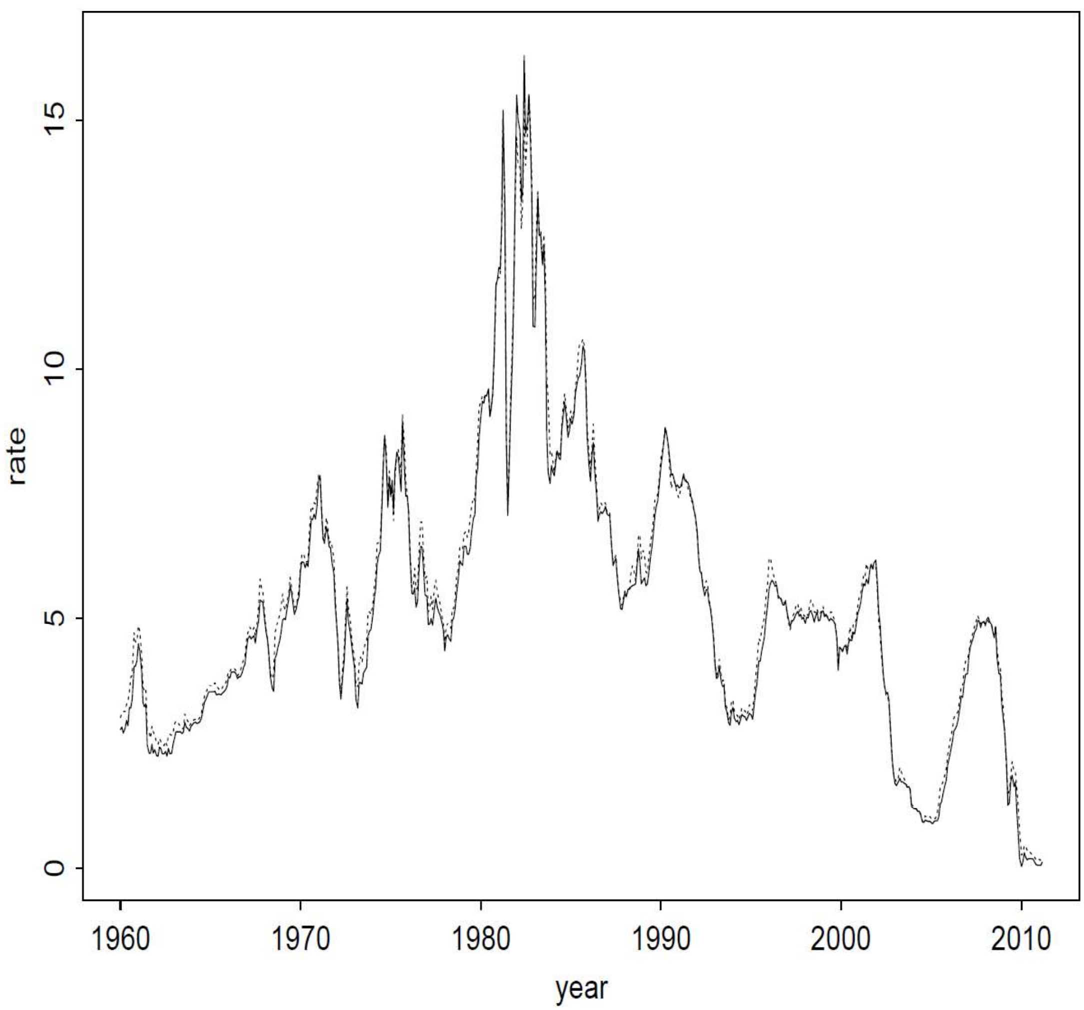

Figure 8: Monthly US interest rates: 3m & 6m TB.

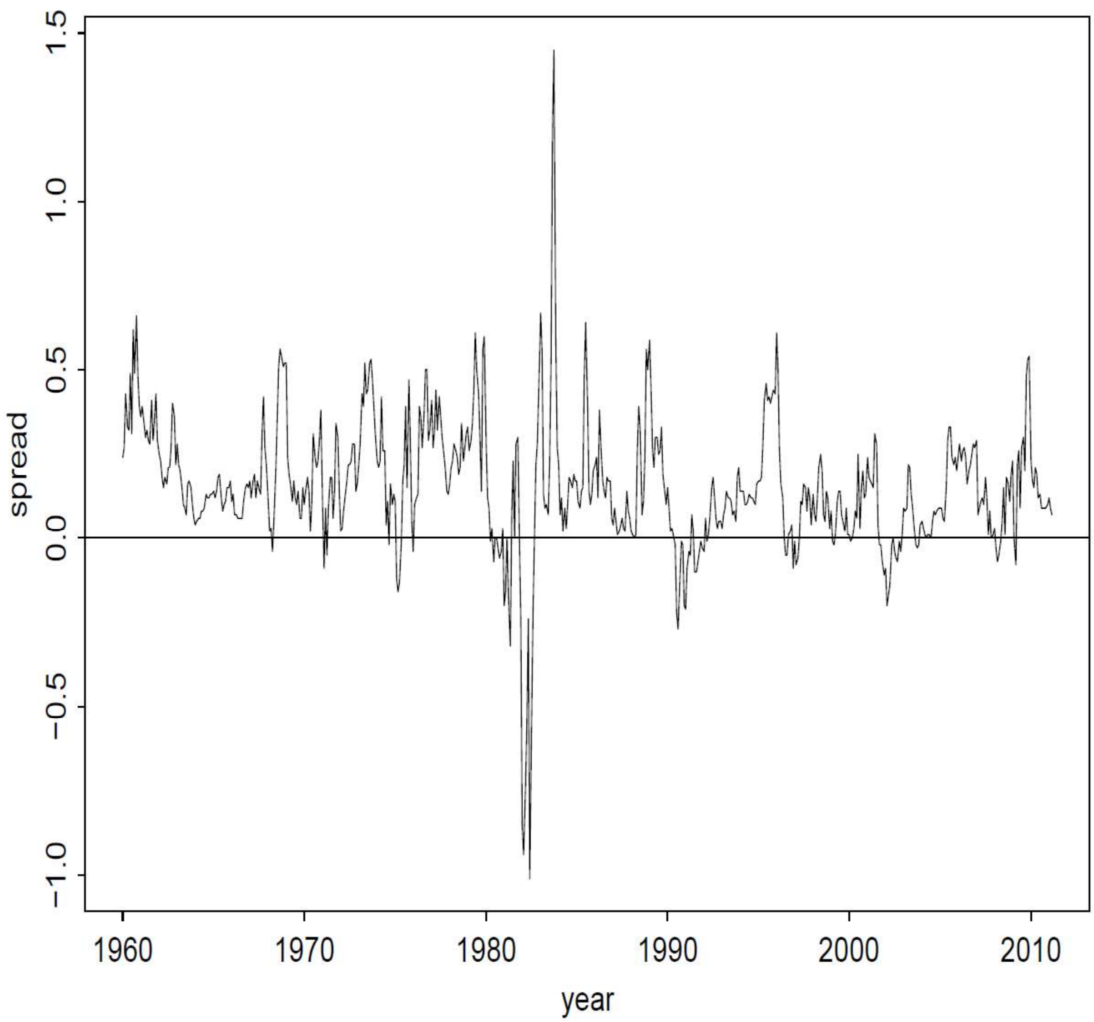

Figure 9: Spread of monthly US interest rates: 3m & 6m TB.

R commands: used to produce plots in Lecture 1.

```
>x=read.table("d-aapl0413.txt",header=T) <== Load Apple stock returns
> dim(x) <== check the size of the data file
[1] 2517 3
> x[1,] <== show the first row of the data
Permno date rtn
1 14593 20040102 -0.004212
> y=ts(x[,3],frequency=252,start=c(2004,1)) <== Create a time-series object in R.
> plot(y,type='l',xlab='year',ylab='rtn')
> title(main='Daily returns of Apple stock: 2004 to 2013')
> par(mfcol=c(2,1)) <== To put two plots on a single page
> y=y*100 <== percentage returns
> hist(y,nclass=50)
> title(main='Percentage returns')
> d1=density(y)
> plot(d1$x,d1$y,xlab='returns',ylab='den',type='l')
> x=read.table("m-tb3ms.txt",header=T) <== Load 3m-TB rates
> dim(x)
[1] 914 4
> y=read.table("m-tb6ms.txt",header=T) <== Load 6m-TB rates
> dim(y)
[1] 615 4
> 914-615
[1] 299
> x[300,] <== Check date of the 3m-TB
year mon day value
300 1958 12 1 2.77
> y[1,] <== Check date of the 1st observation of 6m-TB
year mon day value
1 1958 12 1 3.01
> int=cbind(x[300:914,4],y[,4]) <== Line up the two TB rates
> tdx=(c(1:615)+11)/12+1959
> par(mfcol=c(1,1))
> max(int)
[1] 16.3
> plot(tdx,int[,1],xlab='year',ylab='rate',type='l',ylim=c(0,16.5))
> lines(tdx,int[,2],lty=2) <== Plot the 6m-TB rate on the same frame.
> plot(tdx,int[,2]-int[,1],xlab='year',ylab='spread',type='l')
> abline(h=c(0)) <== Draw a horizontal like to ``zero''.
> x=read.table("q-ko-earns8309.txt",header=T) <== Load KO data
> dim(x)
[1] 107 3
> x[1,]
pends anntime value
1 19830331 19830426 0.0375
> tdx=c(1:107)/12+1983
> plot(tdx,x[,3],xlab='year',ylab='earnings',type='l')
> title(main='EPS of Coca Cola: 1983-2009')
> points(tdx,x[,3])

> y=read.table("d-exuseu.txt",header=T) <== Load USEU exchange rates
> dim(y)
[1] 3567 4
> y[1,]
year mon day value
1 1999 1 4 1.1812
> tdx=c(1:3567)/252+1999
> plot(tdx,y[,4],xlab='year',ylab='eu',type='l')
> title(main='Dollars per Euro')
> r=diff(log(y[,4])) <=== Compute log returns
> plot(tdx[2:3567],r,xlab='year',ylab='rtn',type='l')
> title(main='ln-rtn: US-EU')
> hist(r,nclass=50)
> title(main='useu: ln-rtn')
```

2 Asset Returns

Before discussing different techniques for handling TS, we are going to take a look at returns; what are the various definitions of returns of an asset that we use throughout the course and how we calculate them.

Most financial studies involve returns, instead of prices, of assets. Why?

We prefer to work with returns instead of prices because:

1. Intuitively, returns of an asset are a complete and scale free summary of investment. For example, when you are talking to an average investor and you mention the price of IBM as \$3,000, it's not easy to decipher whether pursuing IBM would be a good investment opportunity. When we speak in terms of returns, i.e say that last month's return on IBM is 1.1%, it would include the price at purchase time as well i.e P_0, which would help with making a decision. In short, returns are easier to communicate with compared to prices.

2. Return series are easier to handle than price series because the former have more attractive statistical properties.

Let P_t be the price of an asset at time t, and assume no dividend.

- One-period simple return: Gross return

$$1 + R_t = \frac{P_t}{P_{t-1}} \quad \text{or} \quad P_t = P_{t-1}(1 + R_t)$$

  Simple net return:

$$R_t = \frac{P_t}{P_{t-1}} - 1$$

- Multiperiod simple return: Gross return

$$1 + R_t(k) = \frac{P_t}{P_{t-k}} = \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times ... \times \frac{P_{t-k+1}}{P_{t-k}} = (1 + R_t)(1 + R_{t-1})...(1 + R_{t-k+1})$$

  The k-period simple net return:

$$R_t(k) = \frac{P_t}{P_{t-k}} - 1$$

  Example: Table below gives five daily closing prices of Apple stock in December 2011.

Date	12/02	12/05	12/06	12/07	12/08	12/09
Price(\$)	389.70	393.01	390.95	389.09	390.66	393.62

  The 1-day gross return of holding the stock from 12/8 to 12/9:

$$1 + R_t = \frac{393.62}{390.66} \approx 1.0076$$

  so that the daily simple return is 0.76%.

  Time interval is important! Default is one year.

- Annualized (average) return: If the asset was held for k years, then the annualized (average) return is defined as

  Annualized $[R_t(k)] = [\Pi_{j=0}^{k-1}(1+R_{t-j})]^{1/k} - 1$
  This is a geometric mean of the k one-period simple gross returns involved.

  Note that

$$[R_t(k)] = [\Pi_{j=0}^{k-1}(1+R_{t-j})]^{1/k}-1 = e^{\log([\Pi_{j=0}^{k-1}(1+R_{t-j})]^{1/k})}-1 = e^{\frac{1}{k}\sum_{j=0}^{k-1}\log(1+R_{t-j})}-1 \approx \frac{1}{k}\sum_{j=0}^{k-1} R_{t-j}$$

  Since arithmetic mean is easier to compute than the geometric mean, this simplifies calculations (assuming that $R_{t-j} \approx 0$).

- Continuously compounding: Before introducing continuously compounded return, we discuss the effect of compounding. Assume that the interest rate of a bank deposit is 10% per annum and the initial deposit is \$1.00. If the bank pays interest once a year, then the net value of the deposit becomes $\$1(1+0.1) = \$1.1$ one year later. If the bank pays interest semi-annually, the 6-month interest rate is $10\%/2 = 5\%$ and the net value is $\$1(1+0.1/2)^2 = \$1.1025$ after the first year. In general, if the bank pays interest m times a year, then the interest rate for each payment is $10\%/m$ and the net value of the deposit becomes $\$1(1+0.1/m)^m$ one year later. Fig 10 gives the results for some commonly used time intervals on a deposit of \$1.00 with interest rate 10% per annum.

Type	#(payment)	Int.	Net
Annual	1	0.1	\$1.10000
Semi-Annual	2	0.05	\$1.10250
Quarterly	4	0.025	\$1.10381
Monthly	12	0.0083	\$1.10471
Weekly	52	$\frac{0.1}{52}$	\$1.10506
Daily	365	$\frac{0.1}{365}$	\$1.10516
Continuously	∞		\$1.10517

Figure 10: The power of compounding with interest rate 10% per annum.

In particular, the net value approaches \$1.1052, which is obtained by exp(0.1) and referred to as the result of continuous compounding. The effect of compounding is clearly seen.

Before moving forward, it is useful to mention the relation $\lim_{m\to\infty}(1+\frac{r}{m})^m = e^r$. Using the previous relation, the net asset value C after n years with continuously compounding return r grows to A:

$$A = C\left(\lim_{m\to\infty}\left(1+\frac{r}{m}\right)^m\right)^n$$
$$\implies A = Ce^{rn}$$

where r is the interest rate per annum, C is the initial capital, and n is the number of years. In other words, the future value of \$1 after n years is e^{rn}, for the current interest rate per annum r. We can write

$$C = Ae^{-rn}$$

which is referred to as the present value of an asset that is worth A dollars n years from now, assuming that the continuously compounded interest rate is r per annum.

■ Continuously compounded (or log) return:

If r_t is the continuously compounded return from time $t-1$ to t, then:

$$P_t = P_{t-1}e^{r_t}$$

$$\Rightarrow r_t = ln(\frac{P_t}{P_{t-1}}) = ln(P_t) - \ln(P_{t-1}) = p_t - p_{t-1}$$

where $p_t = \ln(P_t)$.

We can then obtain the relationship:

$$r_t = ln(\frac{P_t}{P_{t-1}}) = ln(1 + R_t)$$

where R_t is the corresponding simple return.

$$r_t = \ln(1 + R_t) \text{ or } R_t = e^{r_t} - 1$$

If the returns are in percentage, then $r_t = 100 \times \ln(1 + \frac{R_t}{100})$. Therefore $R_t = [e^{\frac{r_t}{100}} - 1] \times 100$.

Multi-period log return:

$$\begin{aligned} r_t(k) &= \ln(1 + R_t(k)) = \ln((1 + R_t)(1 + R_{t-1})...(1 + R_{t-k+1})) \\ &= \ln(1 + R_t) + \ln(1 + R_{t-1}) + ... + \ln(1 + R_{t-k+1}) \\ &= r_t + r_{t-1} + ... + r_{t-k+1} \end{aligned}$$

Example: Consider again the Apple stock price (Fig 10).

1. What is the log return from 12/8 to 12/9?

$$r_t = ln(393.62) - ln(390.66) = 7.5\%.$$

2. What is the log return from day 12/2 to 12/9?

$$r_t(4) = ln((393.62) - ln(389.7) = 1\%$$

Two important formulas:

1. $1 + R_t(k) = (1 + R_t)(1 + R_{t-1})...(1 + R_{t-k+1})$

2. $r_t(k) = r_t + r_{t-1} + ... + r_{t-k+1}$

These two relations are important in practice, e.g. obtain annual returns from monthly returns.

Example: If the monthly log returns of an asset are 4.46%, -7.34% and 10.77%, then what is the corresponding quarterly log return?

$$r = 4.46 - 7.34 + 10.77 = 7.89\%.$$

Example: If the monthly simple returns of an asset are 4.46%, -7.34% and 10.77%, then what is the corresponding quarterly simple return?

$$R = (1 + 0.0446)(1 - 0.0734)(1 + 0.1077) - 1 = 1.0721.1 = 0.0721 = 7.21\%$$

Some advantages of the continuously compounded returns over the simple net returns:

1. the continuously compounded multiperiod return is simply the sum of continuously compounded one-period returns involved.
2. statistical properties of log returns are more tractable.

■ Portfolio return: N assets

$$R_{R,t} = \sum_{i=1}^{N} \omega_i R_{i,t}$$

Example: An investor holds stocks of IBM, Microsoft and CitiGroup. Assume that her capital allocation is 30%, 30% and 40%. Use the monthly simple returns in Table 1.2 of the text. What is the mean simple return of her stock portfolio?

$$E(R_t) = 0.3 \times 1.35 + 0.3 \times 2.62 + 0.4 \times 1.17 = 1.66\%.$$

It is important to note that the continuously compounded returns of a portfolio does not have the above convenient property. It is however a close approximation when the returns are small i.e

$$r_{P,t} \approx \sum_{i=1}^{N} \omega_i r_{i,t} \text{ if } r_{i,t} \approx 0$$

- Dividend payment:

$$R_t = \frac{D_t + P_t}{P_{t-1}} - 1$$

$$r_t = \ln(D_t + P_t) - \ln(P_{t-1})$$

- Excess return: (adjusting for risk)

  In finance, the excess return is thought of as the payoff on an arbitrage portfolio that goes long in an asset and short in the reference asset with no initial investment, an example being borrowing money from the bank to buy IBM stock. It can be formulated as:

$$Z_t = R_t - R_{0t} \text{ and } z_t = r_t - r_{0t}$$

  where r_{0t} denotes the log return of a reference asset (e.g. risk-free interest rate).

Distributional properties of returns

In this section, we dust off some cobwebs and refresh the basic rules of probability. To study asset returns, it is best to begin with their distributional properties. The objective is to understand the behavior of returns across assets overtime. Assume we have N assets over time T. Then $\{r_{it}\}_{t=1}^{T}$ is the *random variable* of interest.

Before we move any further and look more deeply into asset returns, we will delve into some commonly used terminology such as *random variable*, *skewness*, *kurtosis* etc.

Random variables are functions i.e they map the outcomes of a random process to numbers. For a given random process (such as tossing a coin, rolling a pair of dice or predicting the average rainfall in a given area tomorrow), the random variable quantifies the outcomes of the random process. Examples:

1. $X = \begin{cases} 1, & \text{if the coin lands heads} \\ 0,, & \text{if the coin lands tails} \end{cases}$

2. Y=sum of 2 numbers from a die toss.

3. Z=Amount of rain for a given day. $z \in Z \Rightarrow z \in (0, \infty)$(continuous).

4. r= returns from a portfolio. r is continuous since, $r \in (-\infty, \infty)$ i.e the returns can be any value for an interval.

We can assign probabilities to the outcomes of a given random variable. Given a random variable X, its probability distribution function f(x) is a function that assigns a probability to each possible outcome of X. For example, suppose you are flipping a coin; the random variable X is whether the coin shows heads or tails, and to these two outcomes, we assign a probability of Pr(X=heads) = 1/2, and Pr(X=tails) = 1/2.

Continuous variables are those that can take on any value between two numbers. Between 1 and 100, there is an infinite continuum of numbers. Discrete numbers are more like the natural numbers. They take on distinct values. Things that are not normally thought of as numeric can also be coded as discrete numbers. A common example is pregnancy, a variable that is not intrinsically numeric. Pregnancy status might be coded as a zero/one variable, one if the person is pregnant and zero otherwise. Some discrete random variables in economics are: a person's unemployment status, whether two countries are in a monetary union, the number of members in the OPEC, and whether a country was once a colony of the UK. Some discrete random variables in finance include: whether a company is publicly traded or not, the number of times it has offered dividends, or the number of members on the Board.

For a discrete random variable, we call the associated probability distribution function as the probability mass function (pmf). It would assign a probability value to each distinct outcome of a random variable. Naturally, the sum of the probabilities provided for each distinct event should be 1. Examples:

1. $X = \begin{cases} 1, & \text{if the coin lands heads} \\ 0,, & \text{if the coin lands tails} \end{cases}$

   The Probability Mass function would be of the form: $P(X=0) = P(X=1) = \frac{1}{2}$

2. Y=sum of 2 numbers from a die toss. PMF(Y): $P(Y){=}\begin{cases} P(Y=2) = \frac{1}{36}, & P(Y=8) = \frac{5}{36} \\ P(Y=3) = \frac{2}{36}, & P(Y=9) = \frac{4}{36} \\ P(Y=4) = \frac{3}{36}, & P(Y=10) = \frac{3}{36} \\ P(Y=5) = \frac{4}{36}, & P(Y=11) = \frac{2}{36} \\ P(Y=6) = \frac{5}{36}, & P(Y=12) = \frac{1}{36} \\ P(Y=7) = \frac{6}{36} \end{cases}$

When it comes to a continuous random variable, we call the associated probability distribution function as the probability density function (pdf). The pdf measures the "'density"' or the relative likelihood of each outcome. The probability is calculated by taking the integral of the function over some limits i.e

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

where $f(x)$ is the probability density function of X. Some common examples of continuous pdfs are the normal distribution, the student-t distribution, and the chi-squared distribution. Continuous financial random variables include: the percent returns of a stock, the amount of dividends, and the interest rate on bonds. In economics: GDP, the unemployment rate, and the money supply are all continuous variables.For example, assuming that the returns from a portfolio follows a normal standard distribution, its pdf can be written as $f(r) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-1}{2}r^2)$.

Unlike in discrete cases, we have to calculate the probability over an interval rather than a specific point. The values from the probability density function don't necessarily have to add up to 1, but the total area covered by the function in a graph will be 1 (i.e the integral over the function across all possible outcomes).

Finally, we can introduce the cumulative distribution function (CDF) which, in simple terms, calculates the probability of all outcomes below a threshold outcome i.e

$$F(x) = Pr(X \leq x) = \int_{-\infty}^x f(u)du$$

The cumulative distribution function $F(x)$ is an increasing function over a collection of outcomes.

Some important notes on the CDF are:

1. $F(\infty) = 1$.

2. $F(-\infty) = 0$.

3. If $x_1 \leq x_2$, then we have $F(x_1) \leq F(x_2)$.

A very commonly used term associated with the CDF is a quantile. The quantile for a given level of probability is the threshold outcome for which the CDF is equal to or greater than the given probability. For a given probability α,

$$x_\alpha = q(\alpha) = \inf\{x : Pr(X \leq x) \geq \alpha\}$$

For a continuous random variable X, we can write: $x_\alpha = q(\alpha) = F^{-1}(\alpha)$.

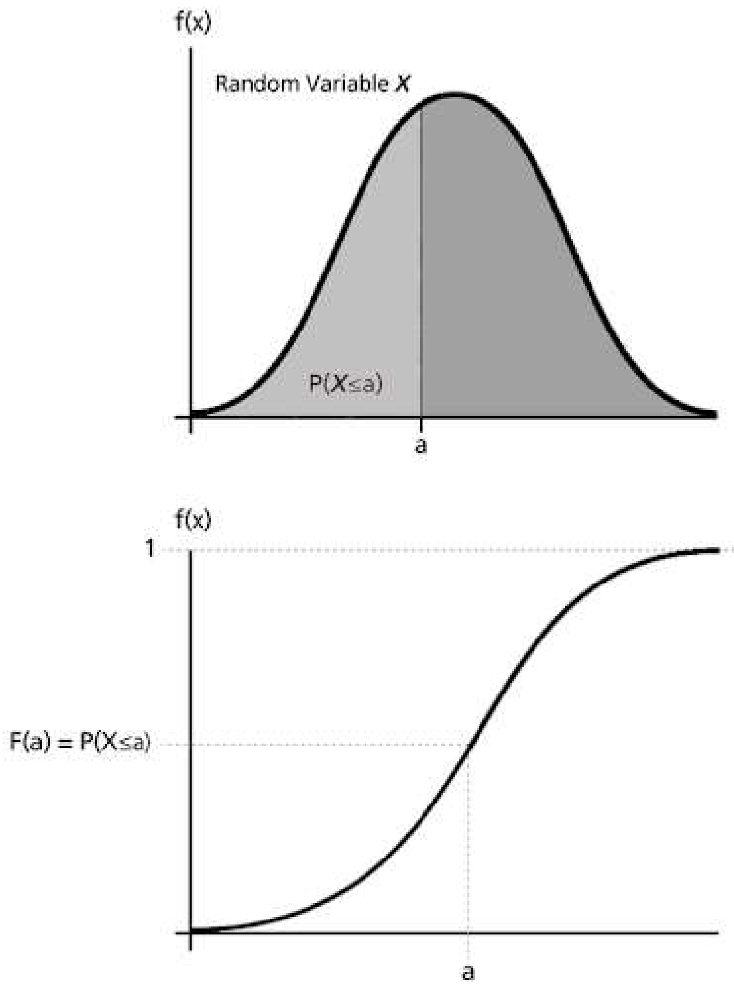

Figure 11: Normal PDF and CDF.

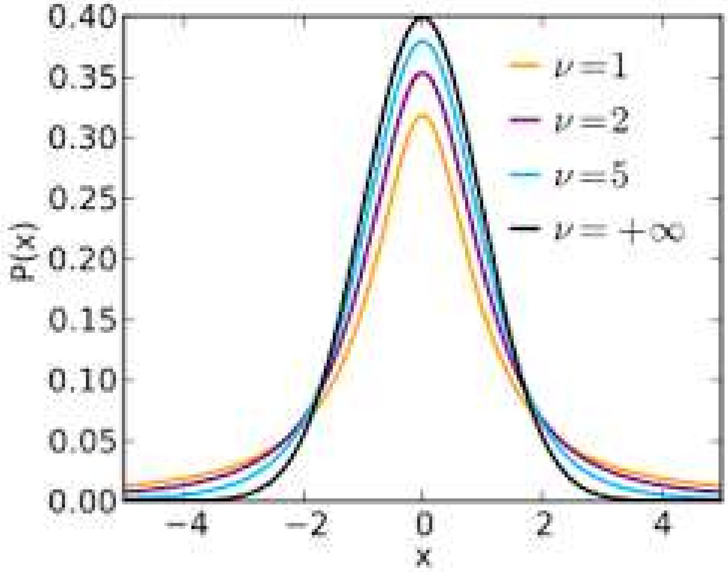

Figure 12: Student-t PDF.

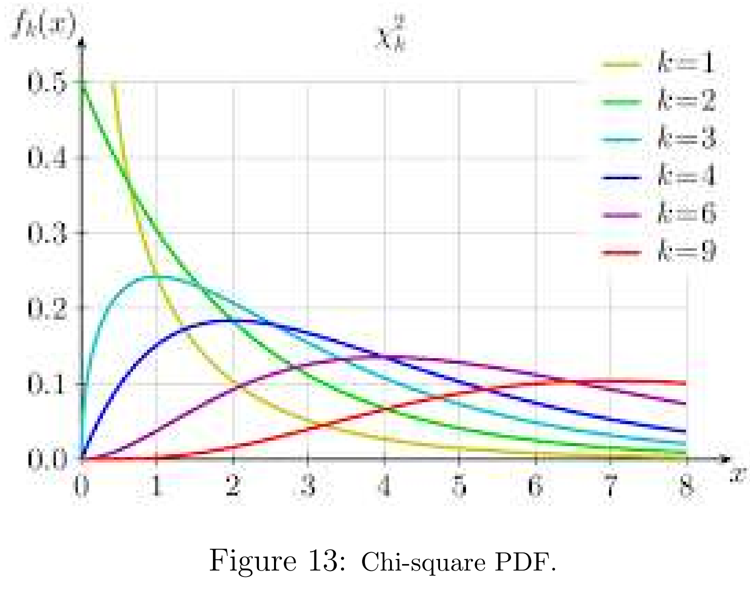

Figure 13: Chi-square PDF.

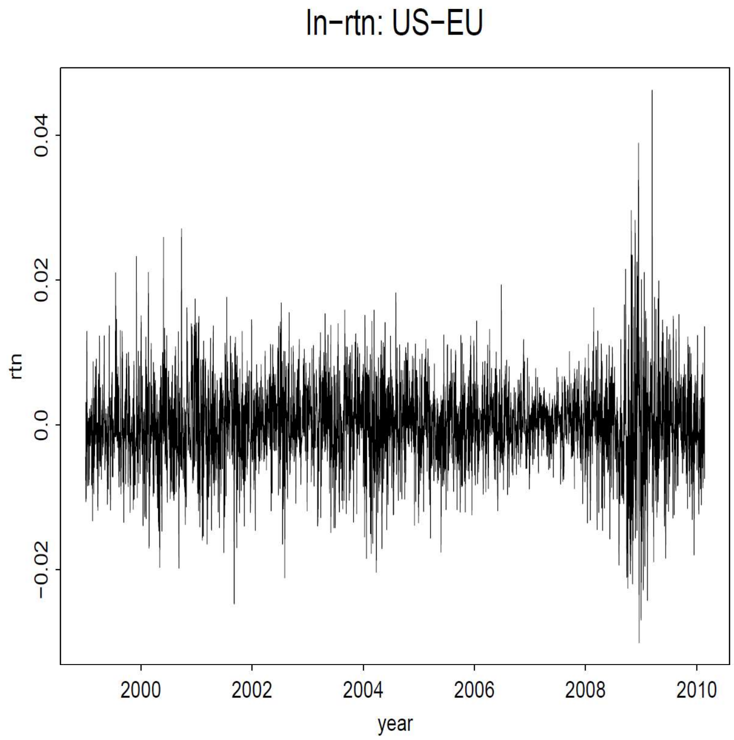

Figure 14: Daily log returns of FX (Dollar vs Euro).

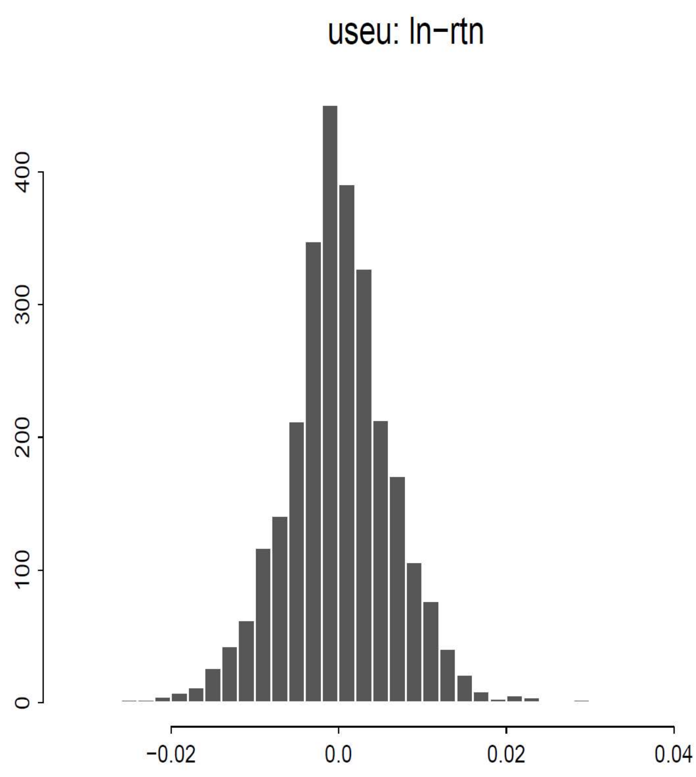

Figure 15: Histogram of daily log returns of FX (Dollar vs Euro).

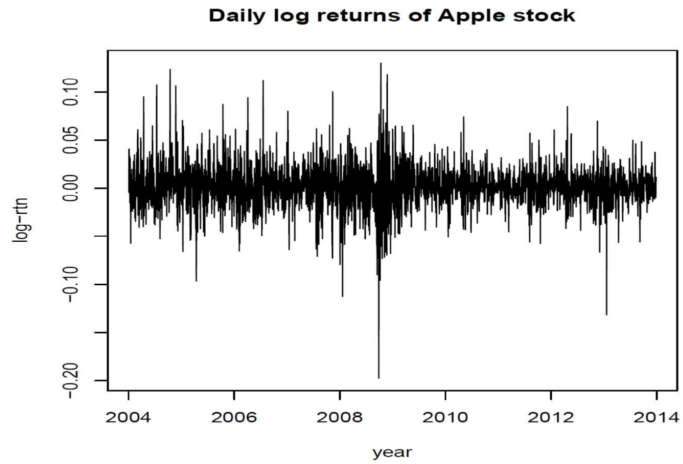

Figure 16: Daily log returns of Apple stock from 2004 to 2013.

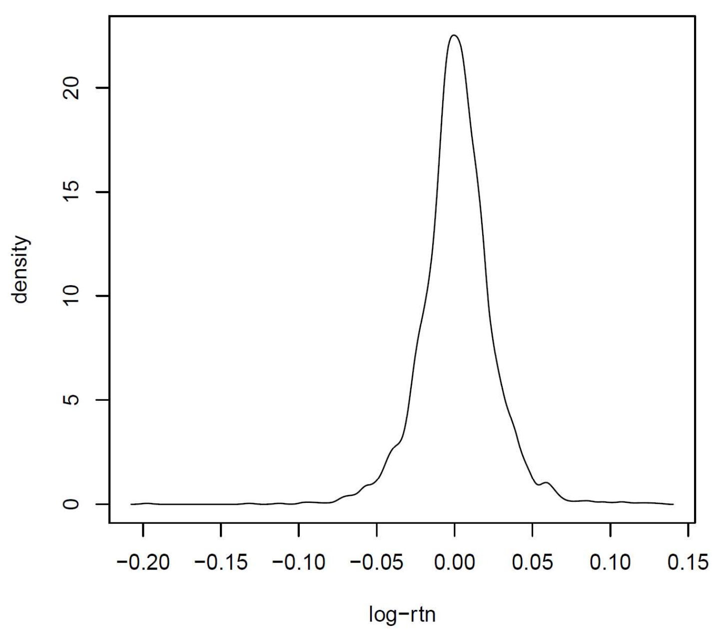

Figure 17: Density of daily log returns of Apple stock: 2004 to 2013.

Some theoretical properties: If X is a continuous variable with density $f(x)$, then

- ℓ-th moment:

  $m'_\ell = E[X^\ell] = \int_{-\infty}^{+\infty} x^\ell f(x)dx$

  First moment: mean or expectation of X. The population mean of $E[X]$ is denoted by μ_X, and the sample mean is represented by $\bar{X}$. $E[X]$ can be understood as the central location of the distribution.

  The first two moments of a random variable ($E[X]$ and $E[X^2]$) uniquely determine a normal density whereas for other distributions, further moments would have to be introduced.

- ℓ-th *central* moment: $m_\ell = E[(X - \mu_X)^\ell] = \int_{-\infty}^{+\infty} (x - \mu_x)^\ell f(x)dx$

  Second central moment is the population variance, $Var(X) = \sigma_X^2 = E[(X - \mu_x)^2]$, which is the average squared deviation of each outcome from the mean.

  Why are the mean and variance of returns important in financial time series? They are concerned with long-term return and risk, respectively.

- The standard deviation is the square root of the variance, Stdev$(X) = \sigma_X = \sqrt{\sigma_X^2}$. The standard deviation σ_X will have the same unit as X, which is why it is commonly used over $Var(X)$, even though they both measure the dispersion of X (i.e to preserve the units).

- The covariance of two random variables X and Y is defined as:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

  It is important to note that covariance is dependent on X and Y. For example, if X is measured in terms of say 10,000kg and Y is measured in terms of 100m , then having a covariance of 10^4 does not mean the strength of their co-movement is 10^4. In order to obtain a value that purely measures the inherent (albeit linear) strength of their co-movement, we would have to normalize the values for X and Y. Normalizing would essentially result in making the inputs without units.

- The correlation between X and Y would then be defined as:

$$\rho_{XY} = Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

  The correlation coefficient ρ is a value between -1 and 1 and is informative regarding the strength of co-movement between X and Y. A perfectly positive correlation implies ρ =1, while perfectly negative correlation would lead to ρ =-1. A completely uncorrelated pair of random variables would have ρ =0.

- Another way to view the correlation coefficient between X and Y is as the covariance between the normalized random variables X^* and Y^*:

$$\rho = cov(X^*, Y^*), \text{ where } X^* = \frac{X - \mu_X}{\sigma_X} \text{ and } Y^* = \frac{Y - \mu_Y}{\sigma_Y}$$

Spearman's rho: Rank correlation. Let $F_x(x)$ and $F_y(y)$ be the cumulative distribution function of X and Y.

$$\rho_s = \rho(F_x(x), F_y(y))$$

That is, the correlation coefficient of probability-transformed variables. It is just the correlation coefficient of the ranks of the data.

Kendall's rho: The nonlinear correlation.

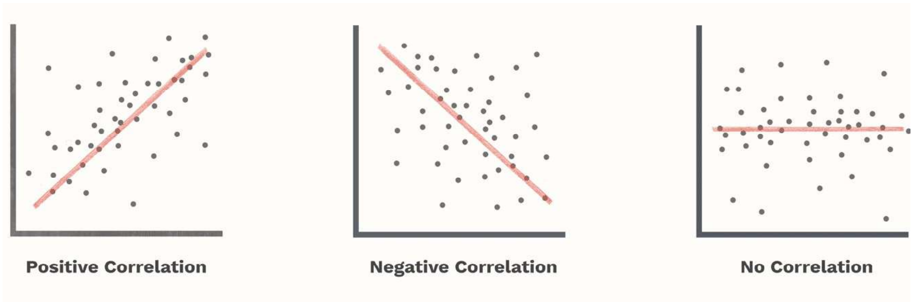

Figure 18: Positive, Negative and Zero linear correlation

- If X and Y are random variables, and a and b are constants, then some simple properties of the statistics listed above are:

  1. $E[a] = a$
  2. $E[aX] = aE[X]$
  3. $\text{Stdev}[aX] = a\text{Stdev}[X]$
  4. $Var[a] = 0$
  5. $Var[aX] = a^2 Var[X]$
  6. $Var[X] = Cov(X, X)$
  7. $Cov(X, Y) = Cov(Y, X)$
  8. $Cov(aX, bY) = abCov(X, Y)$
  9. $Corr(aX, bY) = Corr(X, Y) = Corr(Y, X)$
  10. $E[aX + bY] = E[aX] + E[bY] = aE[X] + bE[Y]$
  11. $Var[aX + bY] = a^2 Var[X] + b^2 Var[Y] + 2abCov(X, Y)$

R Demonstration:

```
> da=read.table("m-ibm6708.txt",header=T)
> head(da)
date ibm sprtn
1 19670331 0.048837 0.039410
.....
6 19670831 -0.013589 -0.011715
> dim(da)
[1] 502 3
> ibm=da$ibm
> sp=da$sprtn
> plot(sp,ibm) <=== not show in this lecture note.
> cor(ibm,sp)
[1] 0.5949041
```

```
> cor(sp,ibm,method="kendall")
[1] 0.4223351
> cor(sp,ibm,method="spearman")
[1] 0.5897587
> x=rank(ibm)
> y=rank(sp)
> cor(x,y)
[1] 0.5897587
> x[1:10]
[1] 375 455 130 408 295 191 453 432 328 305
> y[1:10]
[1] 398 414 44 310 427 157 365 90 219 348
****
> z=rnorm(1000)
> x=exp(z)
> y=exp(20*z)
> cor(x,y)
[1] 0.3187030
> cor(x,y,method='kendall')
[1] 1
> cor(x,y,method='spearman')
[1] 1
```

- Skewness (symmetry):

$$S(X) = E[\frac{(X-\mu_X)^3}{\sigma_X^3}]$$

Symmetry in other words is normalized third moment.

Why is return symmetry of interest in financial study?

Symmetry has important implications in holding short or long financial positions and in risk management.

- Kurtosis (fat-tails):

$$K(X) = E[\frac{(X-\mu_X)^4}{\sigma_X^4}]$$

Kurtosis is the normalized fourth moment. We call $K(X) - 3$ as excess kurtosis, because for a normal distribution $X = N(\mu, \sigma)$, the kurtosis is always 3, which means the excess kurtosis would be 0 for the normal distribution. Why is kurtosis important?

1. Related to volatility forecasting, efficiency in estimation and tests, etc.
2. High kurtosis implies heavy (or long) tails in distribution.

A distribution with positive excess kurtosis ($K(X) > 3$) is said to have heavy tails, implying that the distribution puts more mass on the tails of its support compared to a normal distribution. In practice, a random sample from such a distribution tends to contain more extreme values! Such a distribution is also termed as leptokurtic.

Similarly, a distribution with negative excess kurtosis ($K(X) < 3$) has short tails. Such a distribution is called platykurtic.

- Estimation:

In practice, we do not usually have the required distribution function $f(x)$ for our purposes, and thus we resort to estimating samples from the population instead.

Below we mention some common ways to formulate sample moments from a given sample:

Data: $\{x_1, x_2, ..., x_T\}$

– sample mean:

$$\hat{\mu}_x = \bar{X} = \frac{1}{T}\sum_{t=1}^{T} x_t,$$

– sample variance:

$$\hat{\sigma}_x^2 = \frac{1}{T-1}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^2,$$

– sample skewness:

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^3,$$

– sample kurtosis:

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4}\sum_{t=1}^{T}(x_t - \hat{\mu}_x)^4$$

Under normality assumption, we have

1. $\hat{S}(x) \sim \mathcal{N}(0, \frac{6}{T})$ asymptotically as $T \to \infty$
2. $\hat{K}(x) - 3 \sim \mathcal{N}(0, \frac{24}{T})$ asymptotically as $T \to \infty$

■ Some simple tests for normality (for large T):

1. Test for zero mean: H_0: $\mu = 0$ vs $\mathrm{H}_a : \mu \neq 0$.
   If normality holds, then $t = \frac{\bar{X}}{std(\bar{X})} = \frac{\bar{X}}{\sqrt{\frac{\hat{\sigma}_x^2}{T}}} \sim \mathcal{N}(0,1)$.

   Decision rule: Reject H_0 of zero mean if $|t| > Z_{\alpha/2}$, or p-value is less than α.
2. Test for symmetry: H_0: $S(x) = 0$ vs $\mathrm{H}_a : S(x) \neq 0$.

   If normality holds, then $S^* = \frac{\hat{S}(x)}{\sqrt{\frac{6}{T}}} \sim \mathcal{N}(0,1)$.

   Decision rule: Reject H_o of a symmetric distribution if $|S^*| > Z_{\alpha/2}$, or p-value is less than α.
3. Test for tail thickness: H_0: $K(x) = 3$ vs $\mathrm{H}_a : K(x) \neq 3$.

   If normality holds, then $K^* = \frac{\hat{K}(x)-3}{\sqrt{\frac{24}{T}}} \sim \mathcal{N}(0,1)$.

   Decision rule: Reject H_o of a normal tails if $|K^*| > Z_{\alpha/2}$, or p-value is less than α.
4. A joint test (Jarque-Bera test): H_0: $X \sim$ normal density vs $\mathrm{H}_a : X \nsim$ normal density.

   If normality holds, then $JB = (S^*)^2 + (K^*)^2 \sim \mathcal{X}_2^2$.

   where $\mathcal{X}_2^2$ denotes a chi-squared distribution with 2 degrees of freedom.

   Decision rule: Reject H_o of a normality if $|JB| > \mathcal{X}_2^2(\alpha)$, or p-value is less than α (This is a one sided test).

Recall: Hypothesis testing contains both two sided and one sided tests. As an example, consider we are testing whether the mean of a random variable X, say μ, is not significant i.e $\mu = 0$. Let us choose our level of confidence $\alpha = 0.05$. Then for a standard normal distribution, $z_{0.025}(= z_{\alpha/2}$ for $\alpha = 0.05)$ is the upper 2.5%th quantile.

If $|t| > z_{\alpha/2}$, we reject the null hypothesis $H_0 : \mu = 0$. This is because under H_0, $t \sim N(0, 1)$, therefore having $|t| > z_{\alpha/2}$ is a rare event under our level of confidence. It is useful to note that, for a two sided hypothesis test where the test statistic follows a normal standard distribution, $\alpha = 0.05$ will imply $Z_{\alpha/2} = 1.96$, whereas $\alpha = 0.01$, would lead to $Z_{\alpha/2} = 2.575$.

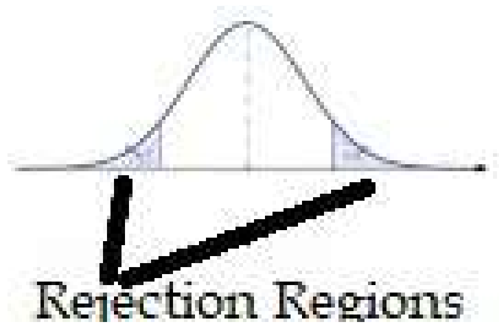

Figure 19: Two-sided test.

We could use p-values to test H_0. If p-value$< \alpha$, we reject H_0. We calculate the p-value as follows:

$$\text{p-value} = 2Pr(z > |t|) = 2\left(1 - Pr(z < |t|)\right)$$

where t is our test statistic.

In a one-sided test:

$$\text{p-value} = Pr(x > |q|)$$

where q is our test statistic. If p-value$< \alpha$, we reject H_0.

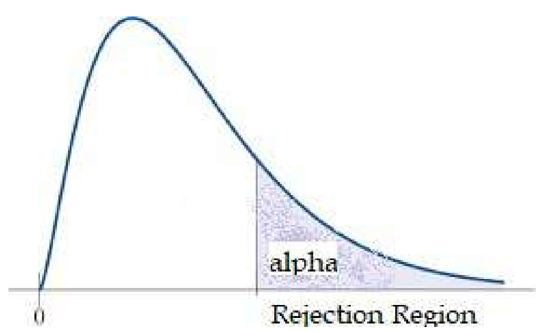

Figure 20: One-sided test.

- Empirical properties of returns:

  Empirical distribution of asset returns tends to be skewed to the left with heavy tails and has a higher peak than normal distribution. See Table 1.2 of the textbook.

  R demonstration: Use monthly IBM stock returns from 1967 to 2008.

```
**** Task: (a) Set the working directory
(b) Load the library ‘‘fBasics’’.
(c) Compute summary (or descriptive) statistics
(d) Perform test for mean return being zero.
(e) Perform normality test using the Jaque-Bera method.
(f) Perform skewness and kurtosis tests.
> setwd("C:/Users/rst/teaching/bs41202/sp2015") <== set working directory
> library(fBasics) <== Load the library ‘‘fBasics’’.
> da=read.table("m-ibm6708.txt",header=T) <=== Load data with header on top
> da[1,] <=== Show the first row of the data
date ibm sprtn
1 19670331 0.048837 0.03941
> ibm=da[,2] <=== Get "ibm" simple returns
> rt=log(ibm+1) <=== Transform into log returns
> plot(rt,type=’l’) <=== Plot log returns with caption.
> title(main=’Time plot of monthly log returns of IBM stock from 1967,3 to 2008.12’)
> mean(rt) <=== Compute sample mean
[1] 0.006208082
> var(rt) <=== Compute sample variance
[1] 0.005258775
> skewness(rt) <=== Compute sample skewness
[1] -0.1353432
> kurtosis(rt) <=== Compute sample excess kurtosis.
[1] 1.693092
> basicStats(rt) <=== Compute all summary (or descriptive) statistics
nobs 502.000000
NAs 0.000000
Minimum -0.303683
Maximum 0.302915
1. Quartile -0.037641
3. Quartile 0.048443
Mean 0.006208
Median 0.005260
Sum 3.116457
SE Mean 0.003237
LCL Mean -0.000151
UCL Mean 0.012567
Variance 0.005259

Stdev 0.072517
Skewness -0.135343
Kurtosis 1.693092
> t.test(rt) <=== Test the mean log return being zero.
One Sample t-test
data: rt
t = 1.9181, df = 501, p-value = 0.05567
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.0001509194 0.0125670842
sample estimates:
mean of x
0.006208082
> normalTest(rt,method=’jb’) <=== Test for the normality assumption
Title:
Jarque - Bera Normalality Test
Test Results:
STATISTIC:
X-squared: 62.8363
P VALUE:
Asymptotic p Value: 2.265e-14
> s3=skewness(rt) <== Perform skewness test
> T=length(rt)
> tst=s3/sqrt(6/T)
> tst
[1] -1.237977
> pv=2*pnorm(tst) <== Compute two-sided p-value of the test statistic.
> pv
[1] 0.2157246
> k4=kurtosis(rt) <== Perform kurtosis test.
> tst=k4/sqrt(24/T)
> tst
[1] 7.743311
```

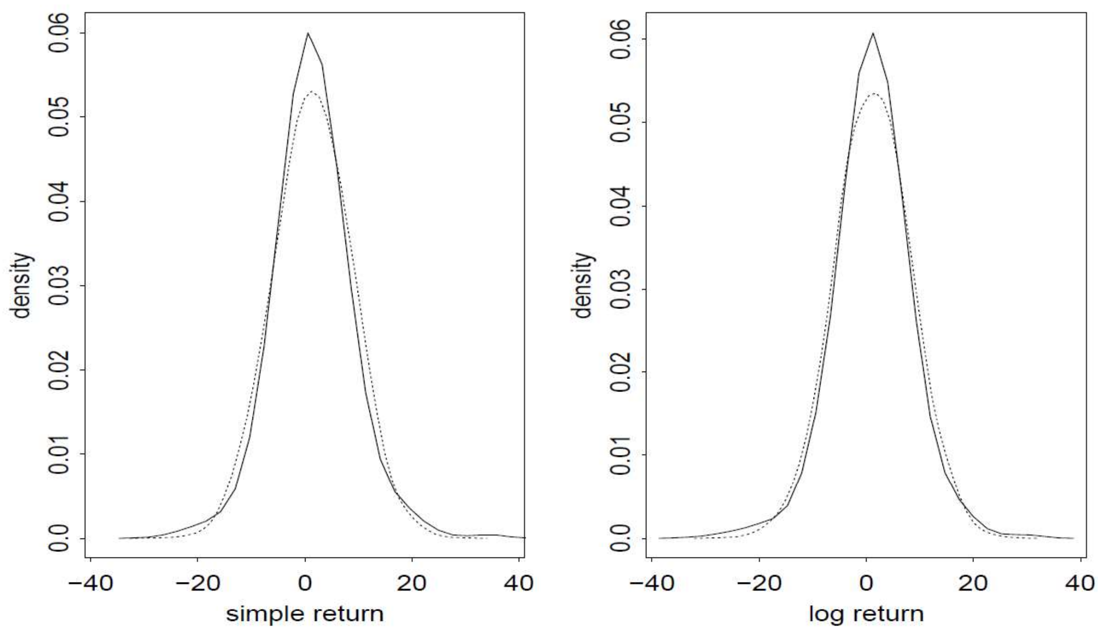

Figure 21: Comparison of empirical IBM return densities (solid) with Normal densities (dashed).

- Normal and lognormal distributions (for self study):

  Y is lognormal if $X = \ln(Y)$ is normal. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = \exp(X)$ is lognormal with mean and variance

$$E[Y] = \exp(\mu + \frac{\sigma^2}{2}), \quad Var[Y] = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$

  Conversely, if Y is lognormal with mean μ_y and variance σ_y^2, then $X = \ln(Y)$ is normal with mean and variance

$$E[X] = \ln[\frac{\mu_y}{1 + \frac{\sigma_y^2}{\mu_y^2}}], \quad Var[X] = \ln[1 + \frac{\sigma_y^2}{\mu_y^2}]$$

  Example: If the log return of an asset is normally distributed with mean 0.0119 and standard deviation 0.0663, then what is the mean and standard deviation of its simple return?

  Answer: Solve this problem in two steps.

  Step 1. Based on the prior results, the mean and variance of $Y_t = \exp(r_t)$ are

$$E[Y] = \exp[0.0119 + \frac{0.0663^2}{2}] = 1.014$$

$$Var[Y] = \exp(2 \times 0.0119 + 0.0663^2)[\exp(0.0663^2) - 1] = 0.0045$$

  Step 2. Simple return is $R_t = \exp(r_t) - 1 = Y_t - 1$. Therefore,

$$E[R] = E(Y) - 1 = 0.014$$

$$Var[R] = Var[Y] = 0.0045$$

  and standard deviation$[R] = \sqrt{0.0045} = 0.067$

- Likelihood function

  The Joint distribution of random variables X and Y is given by $f_{X,Y}(x, y)$. But we can choose to neglect the X and Y terms for notational purposes.

  Basic concept: Joint distribution = Conditional distribution × Marginal distribution, i.e.

$$f(x, y) = f(x|y)f(y)$$

  We will use the Bayes rule to understand the formulation of the likelihood function. In short, the Bayes rule says how we get the conditional probability as below:

$$\underbrace{f(x|y)}_{conditional} = \frac{f(x, y)}{\underbrace{f(y)}_{marginal}} \quad (IMPORTANT!)$$

$$\Longrightarrow f(x, y) = f(x|y)f(y)$$

  where $f(x, y)$ is the joint distribution function on X and Y, $f(x|y)$ is the conditional distribution on X on Y and $f(y)$ is the marginal density function of Y. The above rearrangement is used in Maximum Likelihood Estimation and is also important in the context of Bayesian statistics.

For two consecutive returns r_1 and r_2, we have

$$f(r_1, r_2) = f(r_2|r_1)f(r_1)$$

For three returns r_1, r_2 and r_3, by repeated application,

$$f(r_1, r_2, r_3) = f(r_3|r_1, r_2)f(r_1, r_2) = f(r_3|r_1, r_2)f(r_2|r_1)f(r_1)$$

In general, we have

$$f(r_1, r_2, ..., r_T) = f(r_T|r_1, r_2, ..., r_{T-1})....f(r_2|r_1)f(r_1) = [\Pi_{t=2}^{T} f(r_t|r_1, r_2, ..., r_{t-1})]f(r_1)$$

where $\Pi_{t=2}^{T}$ denotes product.

If $f(r_t|r_1, r_2, ..., r_{t-1})$ is normal with mean μ_t and variance σ_t^2, then likelihood function becomes

$$f(r_1, r_2, ..., r_T) = \left[\Pi_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp(\frac{-(r_t - \mu_t)^2}{2\sigma_t^2})\right] f(r_1).$$

This is the conditional likelihood function of the returns under normality. This is a useful representation since given the available information $\mathcal{F}_{1:t-1} = \{r_1, ...r_{t-1}\}$, we are usually able to get the conditional means (μ_t) (mean modeling) and conditional variances (σ_t) (volatility modeling).

* μ_t: Discussed in Chapter 2

* σ_t^2: Discussed in Chapters 3 and 4.

* Other distributions, e.g. Student-t, can be used to handle heavy tails.

■ Takeaway:

1. Understand the summary statistics of asset returns
2. Understand various definitions of returns and their relationships
3. Learn basic characteristics of FTS
4. Learn the basic R functions.