Lecture 6: Returns and Moments

Cheng Lu

Overview

- Periodic Return
- Moments and Sample Moments
    - Mean and Sample Mean
    - Variance and Sample Variance
    - Moments about the Origin and Central Moments
    - Standardized Moments
        - Skewness
        - Kurtosis
    - Other Moments and Sample Moments
- Homework

Periodic Return

Let P_t be the stock price at time t. Simple return R_t over a period Δt is defined from

Simple Return

$$P_t = P_{t-\Delta t}\left(1 + R_t\right) \Leftrightarrow R_t = \frac{P_t}{P_{t-\Delta t}} - 1$$

n times compounded return over a period Δt is defined from

n Times Compounded Return

$$P_t = P_{t-\Delta t}\left(1 + \frac{R_t^{(n)}}{n}\right)^n \Leftrightarrow R_t^{(n)} = n\left(\left(\frac{P_t}{P_{t-\Delta t}}\right)^{\frac{1}{n}} - 1\right)$$

where n is called compounding frequency, which indicates the number of re-investments between $t - \Delta t$ and t.

Periodic Return

Log return (continuous compounded return) over a period Δt is defined from

Log return

$$P_t = P_{t-\Delta t} \lim_{n\to\infty} \left(1 + \frac{r_t}{n}\right)^n = P_{t-\Delta t} \exp(r_t)$$
$$\Leftrightarrow r_t = \ln \frac{P_t}{P_{t-\Delta t}} = \ln P_t - \ln P_{t-\Delta t}$$

Let the unit period Δt be 1 (day, week, month, year, etc), and t be integer, then we have

$$R_t = \frac{P_t}{P_{t-1}} - 1;\, R_t^{(n)} = n\left(\left(\frac{P_t}{P_{t-1}}\right)^{\frac{1}{n}} - 1\right);\, r_t = \ln \frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1}$$

Periodic Return

Sometimes data may not well ordered

Example

```
> setwd("C:/Users/demonew/Documents/Stevens/Graduate Courses/FE515/20s")
> aapl <- read.csv("AAPL.csv")
> head(aapl) # from 2015-09-11 down to 2013-12-12
        Date   Open   High    Low  Close   Volume Adj.Close
1 2015-09-11 111.79 114.21 111.76 114.21 49441800    114.21
2 2015-09-10 110.27 113.28 109.90 112.57 62675200    112.57
3 2015-09-09 113.76 114.02 109.77 110.15 84344400    110.15
4 2015-09-08 111.75 112.56 110.32 112.31 54114200    112.31
5 2015-09-04 108.97 110.45 108.51 109.27 49963900    109.27
6 2015-09-03 112.49 112.78 110.04 110.37 52906400    110.37
```

Periodic Return

Since date is not well ordered, we need to reverse the rows of the data

Example

```
> aapl <- aapl[nrow(aapl):1,]
> head(aapl)
          Date   Open   High    Low  Close    Volume Adj.Close
440 2013-12-12 562.14 565.34 560.03 560.54  65572500  77.44289
439 2013-12-13 562.85 562.88 553.67 554.43  83205500  76.59874
438 2013-12-16 555.02 562.64 555.01 557.50  70648200  77.02289
437 2013-12-17 555.81 559.44 553.38 554.99  57475600  76.67611
436 2013-12-18 549.70 551.45 538.80 550.77 141465800  76.09309
435 2013-12-19 549.50 550.00 543.73 544.46  80077200  75.22131
```

Periodic Return

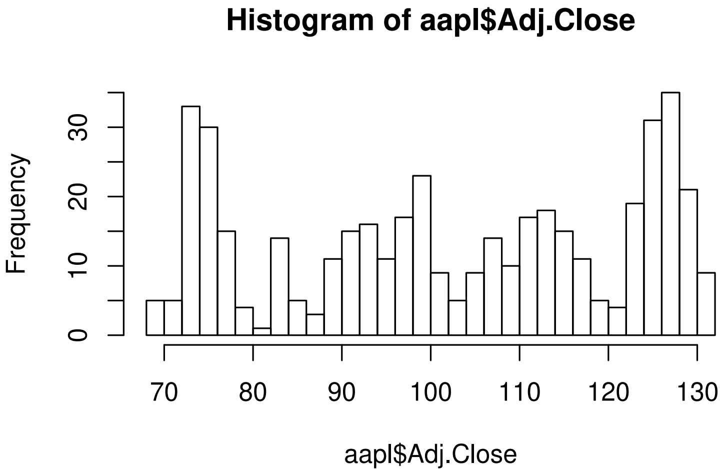
diff() function calculate lagged difference of a vector

Example

```
> x <- 1:10
> diff(x) # x[2]-x[1],x[3]-x[2],...,x[10]-x[9]
[1] 1 1 1 1 1 1 1 1 1
> length(diff(x))
[1] 9
>
> aapl.price <- aapl$Adj.Close # P1,P2,...
> aapl.log.price <- log(aapl.price) # ln(P1),ln(P2),...
> aapl.log.return <- diff(aapl.log.price) # ln(P2)-ln(P1),ln(P3)-ln(P2)...
> aapl.simple.return <- exp(aapl.log.return) - 1 # (P2/P1)-1,(P3/P2)-1,...
>
> hist(aapl$Adj.Close, 40)
> hist(aapl.log.return, 40)
```

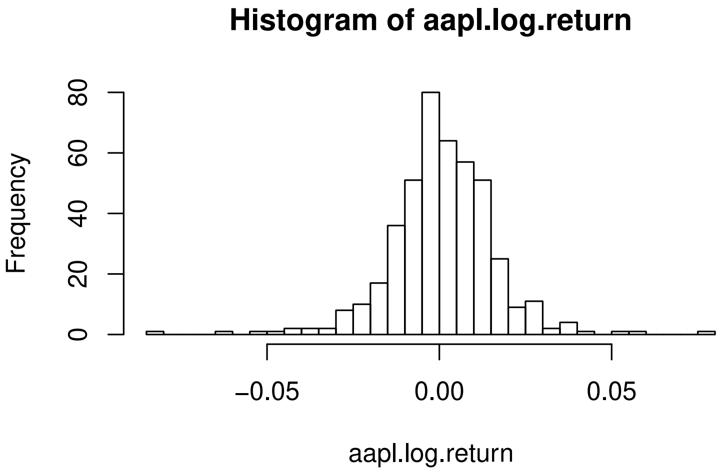
Periodic Return

Histogram of AAPL adjusted close price



Periodic Return

Histogram of AAPL log return is more like normal



Periodic Return

Functions in *quantmod* package calculates periodic returns (default is simple return)

Example

```
> library(quantmod)
> msft <- getSymbols("MSFT", auto.assign = F)
> msft.monthly.return <- periodReturn(msft, period = "monthly")
> msft.annual.return <- periodReturn(msft, period = "yearly")
> msft.all.return <- allReturns(msft)
> head(msft.all.return)
                  daily       weekly monthly quarterly yearly
2007-01-03           NA           NA      NA        NA     NA
2007-01-04 -0.001674548           NA      NA        NA     NA
2007-01-05 -0.005702784 -0.009027115      NA        NA     NA
2007-01-08  0.009784110           NA      NA        NA     NA
2007-01-09  0.001002305           NA      NA        NA     NA
2007-01-10 -0.010013318           NA      NA        NA     NA
```

Mean and Sample Mean

Let X be a random variable, then (population) mean is defined to be the expectation of X

(Population) Mean

$$E[X] = \sum_{i=1}^{k} x_i p_i, \text{for X discrete with probability mass } p_i \text{ at } x_i, i = 1, \ldots, k$$

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx, \text{for X continuous with probability density function } f(x)$$

Suppose $X_1, X_2, \ldots, X_n$ are n observations drawn from population, then sample mean is

Sample Mean

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$$

Mean and Sample Mean

Example

Suppose Y is *one* outcome for rolling a die, then Y takes value $x_i = i$ with probability $p_i = \frac{1}{6}, i = 1, \ldots, 6$, then population mean of Y is given by

$$E[Y] = \sum_{i=1}^{6} x_i p_i = \frac{1}{6} \sum_{i=1}^{6} x_i = 3.5$$

Suppose $Y_1, Y_2, \ldots, Y_n$ are n outcomes for rolling a die, then the sample mean of the observations $Y_1, Y_2, \ldots, Y_n$ is

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

Mean and Sample Mean

Example

Suppose Z follows uniform distribution in $[0,1]$, then population mean of Z is given by

$$E[Z] = \int_{-\infty}^{+\infty} x * 1_{\{0 \leq x \leq 1\}} dx = \int_0^1 x dx = 0.5$$

Suppose $Z_1, Z_2, \ldots, Z_n$ are n observations from uniform distribution in $[0,1]$, then the sample mean of the observations $Z_1, Z_2, \ldots, Z_n$ is

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z_i$$

- Population mean is the expectation of the random variable, which is a number determined from the distribution
- Sample mean is the average of the sample, which is an estimator of the population mean, a random variable

Variance and Sample Variance

(Population) Variance

$$Var(X) = E[(X - E[X])^2] = \sum_{i=1}^{k} (x_i - E[X])^2 \, p_i, \text{for discrete } X$$

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{+\infty} (x - E[X])^2 f(x) dx, \text{for continuous } X$$

Suppose $X_1, X_2, \ldots, X_n$ are n observations, then sample variance is

Sample Variance

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2 \text{ (unadjusted), or } \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2 \text{ (adjusted)}$$

Variance and Sample Variance

Unbiased Estimator

$\hat{\theta}$ is an unbiased estimator of θ when

$$E[\hat{\theta}] = \theta$$

Proposition

Unadjusted sample variance is not an unbiased estimator of population variance

$$E\left[\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2\right] = \frac{n-1}{n}\left(E[X^2] - E^2[X]\right) = \frac{n-1}{n}Var(X)$$

In order to correct the bias, we multiply a factor of $\frac{n}{n-1}$ to the estimator, which gives the adjusted sample variance.

Variance and Sample Variance

Recall sample mean, unadjusted sample variance, and adjusted sample variance

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j,\ m_2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2,\ \hat{m}_2 = \frac{n}{n-1} m_2$$

Since function **mean()** corresponds to the average operator $\frac{1}{n}\sum_{j=1}^{n} \cdot$, we have

Example

```
> x <- rnorm(100000, mean = 0, sd = 5)
> (x.sample.mean <- mean(x)) # sample mean of x
[1] -0.0449756
> (x.sample.variance <- mean((x-x.sample.mean)^2)) # unadjusted sample variance of x
[1] 25.40703
> n <- length(x)
> (x.sample.variance.adjusted <- n/(n-1)*x.sample.variance) # adjusted sample variance of x
[1] 25.40957
> var(x) # built-in function for adjusted sample variance only
[1] 25.40957
> sd(x) # standard deviation
[1] 5.04079
```

Moments about the Origin and Central Moments

r-th moment about the origin

$$\mu_r' = E\left[X^r\right]$$

r-th sample moment about the origin

$$m_r' = \frac{1}{n}\sum_{j=1}^{n} X_j^r$$

r-th central moment

$$\mu_r = E\left[(X - E\left[X\right])^r\right]$$

r-th sample central moment

$$m_r = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^r, \text{(unadjusted)}$$

Moments about the Origin and Central Moments

- (Population) Mean is the first moment about the origin
- Sample mean is the first sample moment about the origin
- (Population) Variance is the second central moment
- Sample variance is the second sample central moment
- Sample moments (sample mean, sample variance, etc) are estimators of population moments (population mean, population variance, etc)
- When people say "moments" of a random variable or a known distribution, they usually mean population moments.
- When people say "moments" of a vector of observations, they usually mean sample moments.
- Softwares (e.g. R) usually calculates sample moments

Moments about the Origin and Central Moments

We can calculate higher sample moments ourselves, since

$$m'_r = \frac{1}{n}\sum_{j=1}^{n} X_j^r,\, m_r = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})^r$$

and the function **mean()** corresponds to $\frac{1}{n}\sum_{j=1}^{n}\cdot$ in R, then for $r = 3$ or $r = 4$,

Example

```
> (m3.prime <- mean(x^3))# 3rd sample moment about the origin
[1] -8.607392
> (m4.prime <- mean(x^4))# 4th sample moment about the origin
[1] 1913.961
> (m3 <- mean((x-mean(x))^3))# 3rd sample central moment
[1] -5.179213
> (m4 <- mean((x-mean(x))^4))# 4th sample central moment
[1] 1912.721
```

Standardized Moments

Standardized Moments

The rth standardized moment of X is defined to be

$$\tilde{\mu}_r = \frac{\mu_r}{\sigma^r} = \frac{\mu_r}{(\mu_2)^{\frac{r}{2}}}$$

where μ_r is the rth central moment of X, σ is the standard deviation of X.

Sample Standardized Moments

The rth sample standardized moment of observations $X_1, X_2, \ldots, X_n$ is defined to be

$$\tilde{m}_r = \frac{m_r}{(m_2)^{\frac{r}{2}}}$$

where m_r is the rth sample central moment of $X_1, X_2, \ldots, X_n$.

Skewness

Skewness

Skewness of X is defined to be the 3rd standardized moment $\tilde{\mu}_3$ of X.

Sample Skewness

(Unadjusted) Sample skewness of $X_1, X_2, \ldots, X_n$ is the 3rd sample standardized moment $\tilde{m}_3$. Adjusted sample skewness is defined to be

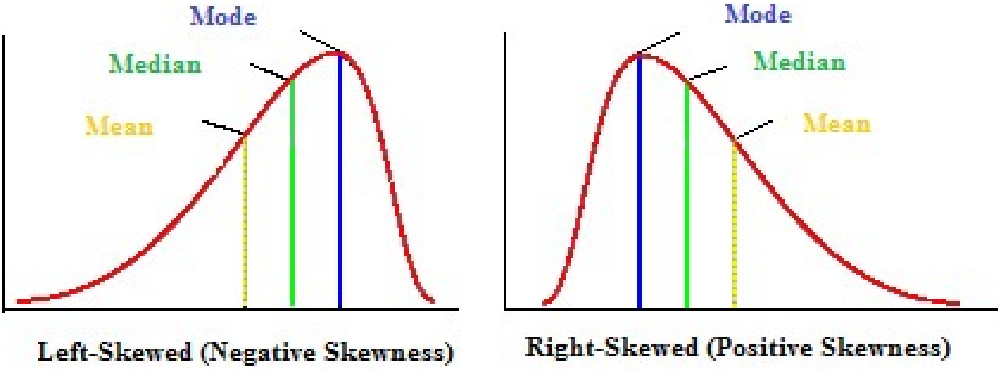
$$\hat{m}_3 = \frac{\sqrt{n(n-1)}}{n-2}\tilde{m}_3$$

.

- For normal distribution, skewness is 0
- When skewness is negative, the distribution is said to be left skewed
- When skewness is positive, the distribution is said to be right skewed

Skewness

In probability theory and statistics, skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.



Kurtosis

Kurtosis

Kurtosis of X is defined to be the 4th standardized moment $\tilde{\mu}_4$ of X.

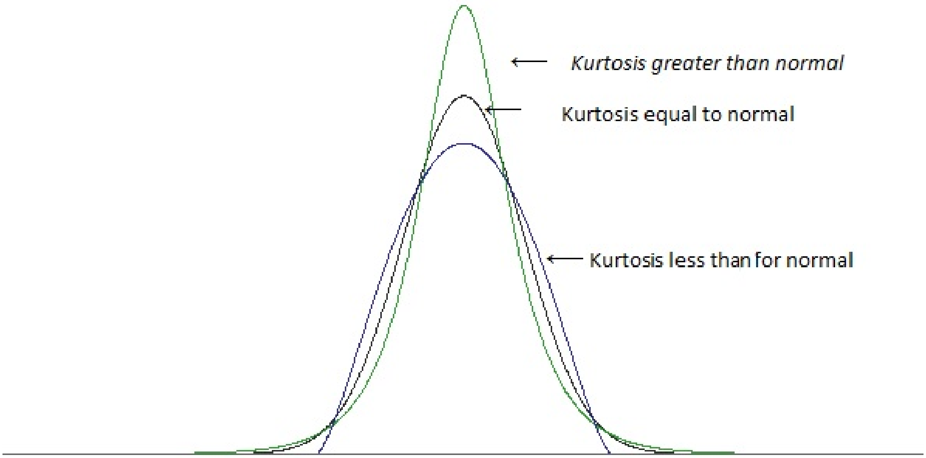
Sample Kurtosis

(Unadjusted) Sample kurtosis of $X_1, X_2, \ldots, X_n$ is the 4th sample standardized moment $\tilde{m}_4$.
Adjusted sample kurtosis is defined to be

$$\hat{m}_4 = \frac{n-1}{(n-2)(n-3)}((n+1)\tilde{m}_4 - 3(n-1)) + 3$$

- For normal distribution, kurtosis is 3
- Excess kurtosis $=$ kurtosis $-$ 3 (Some packages use excess kurtosis to define kurtosis)
- If kurtosis > 3, the distribution has fat tail (Leptokurtic)
- If kurtosis < 3, the distribution has thin tail (Platykurtic)

Kurtosis


Kurtosis greater than normal
Kurtosis equal to normal
Kurtosis less than for normal

Other Moments and Sample Moments

Variance

$$Var(X) = E[(X - E[X])^2] = E[X^2] - E^2[X]$$

Covariance

$$Cov(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

Sample Covariance

$$\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(Y_j - \bar{Y})$$

Covariance is also a central cross-moment.