Lecture 1. Returns and Their Statistical Properties

1. Asset Return

Definition 1.1. Let P_t be the price of an asset at time t. The single-period return from time $t-1$ to time t, denoted by R_t, is defined as

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

and the single-period gross return from time $t-1$ to time t is given by

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

- The multi-period return from time $t-k$ to time t, denoted by $R_{t-k\to t}$, is

$$\begin{aligned} R_{t-k\to t} &= \frac{P_t}{P_{t-k}} - 1 \\ &= \frac{P_{t-k+1}}{P_{t-k}} \times \frac{P_{t-k+2}}{P_{t-k+1}} \times \cdots \times \frac{P_t}{P_{t-1}} - 1 \\ &= (1+R_{t-k+1}) \times (1+R_{t-k+2}) \times \cdots \times (1+R_t) - 1, \end{aligned} \tag{1.1}$$

  which is a product of k single-period gross returns minus one.

Definition 1.2. Let p_t be the logarithm of P_t. The single-period log return from time $t-1$ to time t, denoted by r_t, is defined as

$$r_t = \ln(1+R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) = p_t - p_{t-1}.$$

- From (1.1), the multi-period log return from time $t-k$ to time t, denoted by $r_{t-k\to t}$, is given by

$$\begin{aligned} r_{t-k\to t} &= \ln(1+R_{t-k\to t}) \\ &= \ln\left((1+R_{t-k+1}) \times (1+R_{t-k+2}) \times \cdots \times (1+R_t)\right) \\ &= \ln(1+R_{t-k+1}) + \ln(1+R_{t-k+2}) + \cdots + \ln(1+R_t) \\ &= r_{t-k+1} + r_{t-k+2} + \cdots + r_t, \end{aligned}$$

  which is a sum of k single-period log returns.

Remark 1.3. Let r_t be the multiple compound interest rate with m being the number of compounding periods from time $t-1$ to time t. Then we have

$$1+R_t = \left(1+\frac{r_t}{m}\right)^m. \tag{1.2}$$

As $m \to \infty$, the RHS of (1.2) converges to

$$\lim_{m\to\infty}\left(1+\frac{r_t}{m}\right)^m = e^{r_t}.$$

Therefore, we know $1+R_t = e^{r_t}$ in the continuous compounding case, thereby meaning that the log return is equivalent to the continuously compound return, i.e., $r_t = \ln(1+R_t)$.

Remark 1.4. A first-order Talyor approximation of $f(x) = \ln(1+x)$ at $x = x_0$ is

$$\begin{aligned} f(x)|_{x=x_0} &\approx f(x_0) + f(x_0)^{'}(x-x_0) \\ &\approx \ln(1+x_0) + \frac{x-x_0}{1+x_0}. \end{aligned}$$

When x_0 is close to zero around the origin, we have

$$\ln(1+x) \approx x,$$

which implies that the log return, $\ln(1+R_t)$, well approximates the return, R_t, since returns are typically close to zero.

2. Statistical Properties of Returns

2.1. Finite and Asymptotic Distributions

Definition 2.1. The probability density function (pdf) of a gamma(α,β) distribution is defined as

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$$

for $0 < x < \infty$, $\alpha > 0$, and $\beta > 0$, where the gamma function $\Gamma(\alpha)$ is given by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt.$$

Definition 2.2. The pdf of a chi-squared distribution with p degrees of freedom is defined as

$$f(x) = \frac{1}{\Gamma(p/2)2^{p/2}} x^{(p/2)-1} e^{-x/2}$$

for $0 < x < \infty$. A chi-squared distribution is a special case of a gamma distribution when $\alpha = p/2$ and $\beta = 2$.

Definition 2.3. The pdf of a normal distribution with mean μ and variance σ^2 is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

for $-\infty < x < \infty$.

- If $X \sim N(\mu, \sigma^2)$, it shows
$$Z = \frac{X-\mu}{\sigma} \sim N(0,1),$$
meaning that the random variable Z has a standard normal distribution.

Definition 2.4. If X is a random variable whose logarithm is normally distributed, i.e., $\log(X) \sim N(\mu, \sigma^2)$, then X has a lognormal distribution of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{x} \exp\left(-\frac{(\log x-\mu)^2}{2\sigma^2}\right)$$

for $0 < x < \infty$.

Example 2.5. Suppose that the log return r_t follows a normal distribution with mean μ and variance σ^2, i.e., $r_t \sim N(\mu, \sigma^2)$. This means that the gross return, $1+R_t$, follows a lognormal distribution because $r_t = \ln(1+R_t)$.

Theorem 2.6. *Let $X_1, \ldots, X_n$ be an iid random sample from a $N(\mu, \sigma^2)$ distribution and let $\bar{X} = (1/n)\sum_{i=1}^n X_i$. Then it shows*

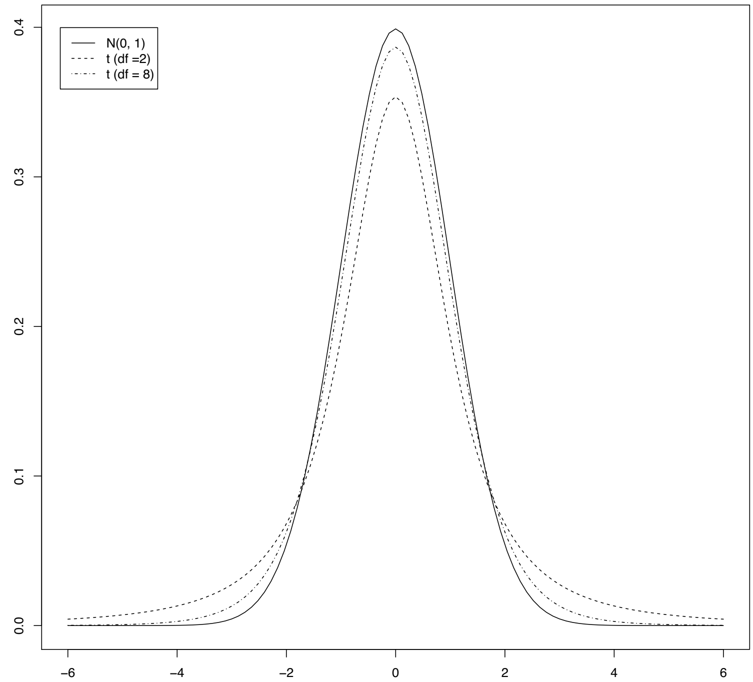
$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Definition 2.7. Let $X_1, \ldots, X_n$ be an iid random sample from a $N(\mu, \sigma^2)$ distribution. The quantity

$$\frac{\bar{X}-\mu}{s/\sqrt{n}}$$

has a Student t distribution with $n-1$ degrees of freedom.

Example 2.8. The t-distribution has a fat tail. As the degrees of freedom increase, the tail becomes less fat.



Definition 2.9. Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid random variables and let $F_n(x)$ be the cumulative density function (cdf) of X_n. The sequence X_n is said to converge in distribution to X if there exists a cdf $F(x)$ such that

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at any value x at which $F(\cdot)$ is continuous. This is indicated as

$$X_n \approx X.$$

Theorem 2.10. *Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of iid random variables with mean μ and variance σ^2. The central limit theorem (CLT) refers to the case that, as n increases, a sequence of iid random variables $\sqrt{n}(\overline{X}-\mu)$ converges in distribution to a Gaussian random variable with mean zero and variance σ^2, i.e.,*

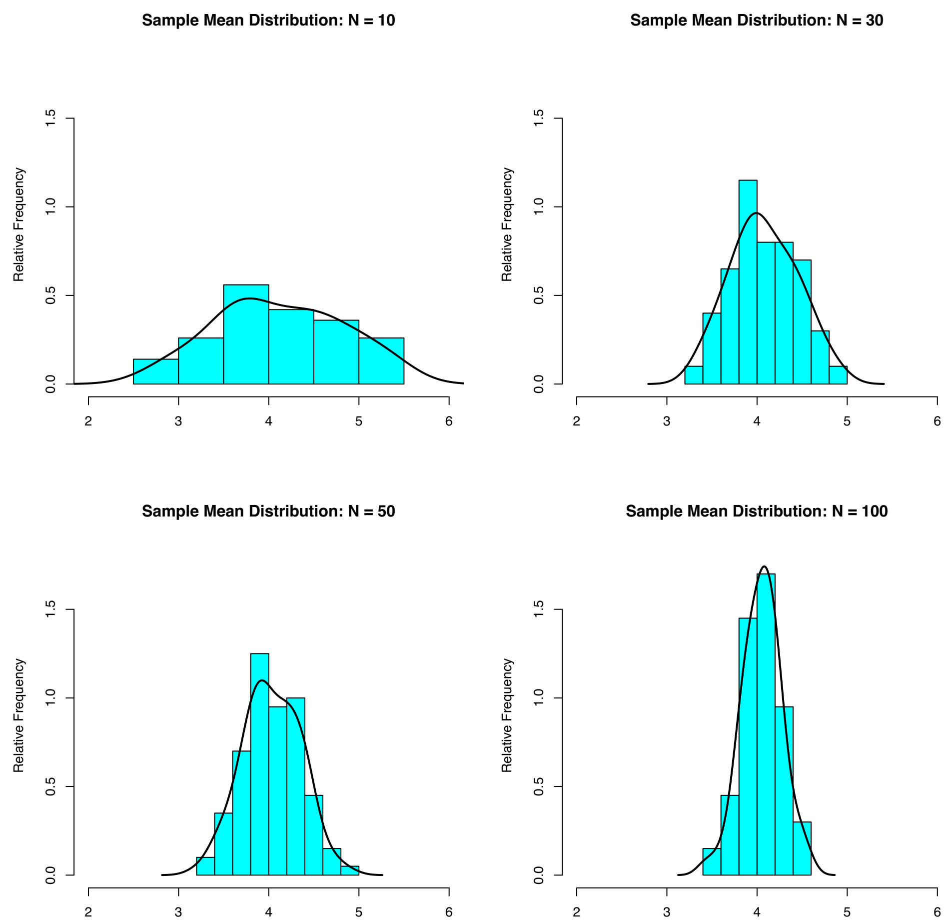
$$\sqrt{n}(\overline{X}-\mu) \approx N(0,\sigma^2)$$

or equivalently

$$\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right).$$

- The CLT implies when n is sufficiently large, the sample mean $\bar{X}$ is approximately distributed as a normal distribution of mean μ and variance σ^2/n.

Example 2.11. Suppose that a population is governed by a uniform distribution U(0,8). The sample mean is computed from samples of size 10, 30, 50, and 100 respectively.



2.2. Population Moments

Definition 2.12. The kth moment of a random variable X is given by

$$m_k' = E[X^k].$$

- The first moment, denoted by μ, is called the mean or expectation of X. It measures the central location of a distribution.

Definition 2.13. The kth central moment of a random variable X is given by

$$m_k = E[(X-\mu)^k].$$

- The second central moment, denoted by σ^2, is called the variance of X. The positive square root, σ, of the variance is the standard deviation of X. The variance (or standard deviation) measures the variability or uncertainty of X.

Definition 2.14. The skewness and kurtosis of X are given by

$$\begin{aligned} S(X) &= E\left[\frac{(X-\mu)^3}{\sigma^3}\right] \\ K(X) &= E\left[\frac{(X-\mu)^4}{\sigma^4}\right]. \end{aligned}$$

- The skewness measures the symmetry of X with respect to its mean. A symmetric distribution has $S(X)=0$, while an asymmetric distribution has $S(X)\neq 0$.

- The quantity $K(X)-3$ is called excess kurtosis. The excess kurtosis of a normal random variable is zero. A distribution with positive excess kurtosis is said to have heavy tails, meaning that the distribution puts more mass on tails than a normal distribution does. Such a distribution is said to be leptokurtic.

Theorem 2.15. *Let $\{x_1,\ldots,x_T\}$ be an iid random sample of X with T observations. Under $H_0 : S(X)=0$, it shows*

$$\frac{\hat{S}(X)}{\sqrt{6/T}} \approx N(0,1),$$

where

$$\begin{aligned} \hat{S}(X) &= \frac{1}{(T-1)\hat{\sigma}^3}\sum_{t=1}^{T}(x_t-\hat{\mu})^3 \\ \hat{\mu} &= \frac{1}{T}\sum_{t=1}^{T}x_t \\ \hat{\sigma}^2 &= \frac{1}{T-1}\sum_{t=1}^{T}(x_t-\hat{\mu})^2. \end{aligned}$$

Remark 2.16. H_0 is rejected at the 5% level if p-value is less than 0.05.

Theorem 2.17. *Let $\{x_1, \ldots, x_T\}$ be a random sample of X with T observations. Under H_0 : $K(X) - 3 = 0$, it shows*

$$\frac{\hat{K}(X) - 3}{\sqrt{24/T}} \approx N(0,1),$$

where

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

Theorem 2.18. *Let $\{x_1, \ldots, x_T\}$ be a random sample of X with T observations. Under the null hypothesis that X is normally distributed, it shows*

$$JB \approx \chi^2_{(2)},$$

where the Jarque-Bera test statistic is computed as

$$JB = \frac{N}{6} \left(\hat{S}^2(X) + \frac{(\hat{K}(X) - 3)^2}{4} \right).$$

Remark 2.19. We might assume that the return R_t follows a normal distribution. The normality assumption induces several difficulties, however. First, the lower bound of R_t is -1, while a normal distribution has no lower bound. Second, the multi-period return is not normally distributed since it is a product of single-period gross returns. Third, asset returns often exhibit positive excess kurtosis.

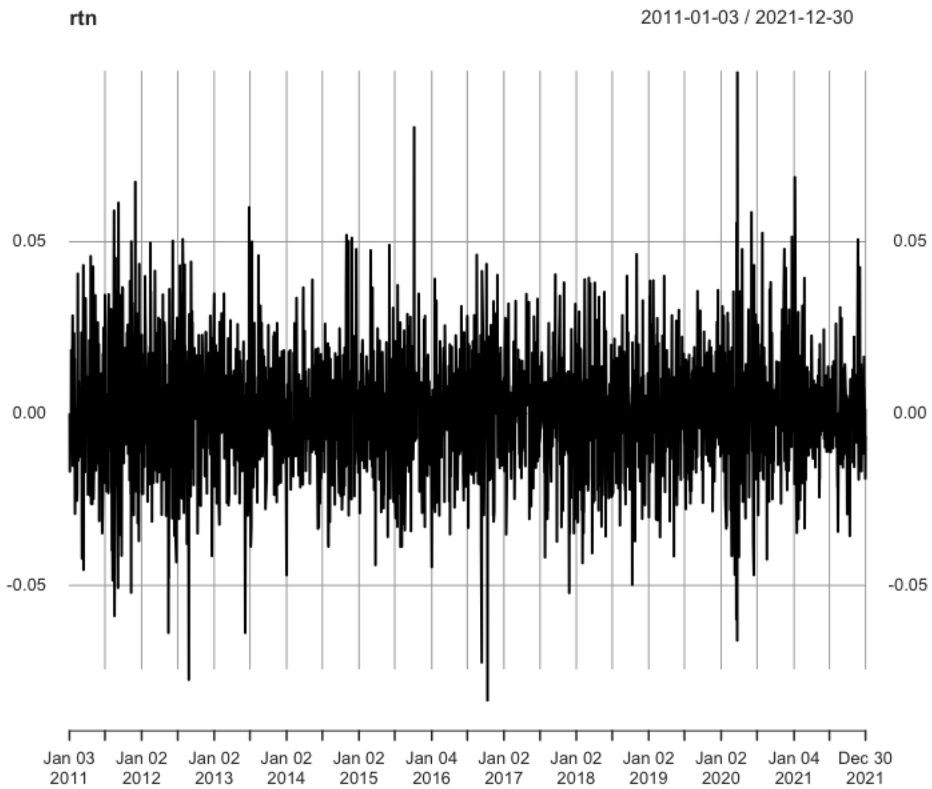
Remark 2.20. It is better to assume that the log return r_t follows a normal distribution, which has two advantages. First, r_t has no lower bound. Second, the multi-period log return is also normally distributed since it is a sum of single-period log returns. But, this approach is still problematic in that the positive excess kurtosis is not addressed.

3. R Code

Example 3.1. Daily adjusted closing prices of the Samsung Electronics stock over the period from 2011/01/03 to 2021/12/30

```
> mydat <- read.csv("data1_1.csv", header = T)
> head(mydat)
    TRD_DD ADJ_PRC
1 20110103   19160
2 20110104   19160
3 20110105   18840
```

```
4 20110106    18600
5 20110107    18420
6 20110110    18340
> library(xts)
> library(data.table)
> library(fBasics)
> prc <- xts(mydat$ADJ_PRC,
             order.by = as.Date(as.character(mydat$TRD_DD), "%Y%m%d"))
> rtn <- diff(log(prc))
> plot(rtn)
```




```
> mydat <- cbind(mydat,
                 ADJ_PRC1 = shift(mydat$ADJ_PRC, n = 1, type = "lag"))
> names(mydat)[2] <- "ADJ_PRC2"
> mydat <- cbind(mydat,
                 RTN = (mydat$ADJ_PRC2-mydat$ADJ_PRC1)/mydat$ADJ_PRC1)
> head(mydat)
    TRD_DD ADJ_PRC2 ADJ_PRC1          RTN
```

```
1 20110103    19160       NA           NA
2 20110104    19160    19160  0.000000000
3 20110105    18840    19160 -0.016701461
4 20110106    18600    18840 -0.012738854
5 20110107    18420    18600 -0.009677419
6 20110110    18340    18420 -0.004343105
> (S <- skewness(mydat$RTN, na.rm = T))
[1] 0.1877175
> n <- length(mydat$RTN)-1
> teststat <- S/sqrt(6/n)
> 2*(1-pnorm(teststat))
[1] 6.622482e-05
```

- We reject $H_0 : S(X) = 0$ at the 5% level, meaning that daily returns exhibit an asymmetric distribution.

```
> (K <- kurtosis(mydat$RTN, na.rm = T))
[1] 1.8058
> teststat <- K/sqrt(24/n)
> 2*(1-pnorm(t.stat))
[1] 0
```

- We reject $H_0 : K(X) = 3$ at the 5% level, meaning that daily returns have heavy tails.

```
> normalTest(mydat$RTN, method = "jb", na.rm = T)
Test Results:
STATISTIC:
X-squared:  385.5937
P VALUE:
Asymptotic p Value:  < 2.2e-16
```

- The normality assumption is rejected at the 5% level.