

Returns

2.1 Introduction

The goal of investing is, of course, to make a profit. The revenue from investing, or the loss in the case of a negative revenue, depends upon both the change in prices and the amounts of the assets being held. Investors are interested in revenues that are high relative to the size of the initial investments. Returns measure this, because returns on an asset, e.g., a stock, a bond, a portfolio of stocks and bonds, are changes in price expressed as a fraction of the initial price.

2.1.1 Net Returns

Let P_t be the price of an asset at time t . Assuming no dividends, the *net return* over the holding period from time $t - 1$ to time t is

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

The numerator $P_t - P_{t-1}$ is the revenue or profit during the holding period, with a negative profit meaning a loss. The denominator, P_{t-1} , was the initial investment at the start of the holding period. Therefore, the net return can be viewed as the relative revenue or profit rate.

The revenue from holding an asset is

$$\text{revenue} = \text{initial investment} \times \text{net return}.$$

For example, an initial investment of \$10,000 and a net return of 6% earns a revenue of \$600. Because $P_t \geq 0$,

$$R_t \geq -1, \tag{2.1}$$

so the worst possible return is -1 , that is, a 100% loss, and occurs if the asset becomes worthless.

2.1.2 Gross Returns

The simple *gross return* is

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

For example, if $P_t = 2$ and $P_{t+1} = 2.1$, then $1 + R_{t+1} = 1.05$, or 105%, and $R_{t+1} = 0.05$, or 5%.

Returns are scale-free, meaning that they do not depend on units (dollars, cents, etc.). Returns are *not* unitless. Their unit is time; they depend on the units of t (hour, day, etc.). In the example, if t is measured in years, then, stated more precisely, this net return is 5% per year.

The *gross return over the most recent k periods* is the product of the k single-period gross returns (from time $t - k$ to time t):

$$\begin{aligned} 1 + R_t(k) &= \frac{P_t}{P_{t-k}} = \left(\frac{P_t}{P_{t-1}} \right) \left(\frac{P_{t-1}}{P_{t-2}} \right) \cdots \left(\frac{P_{t-k+1}}{P_{t-k}} \right) \\ &= (1 + R_t) \cdots (1 + R_{t-k+1}). \end{aligned}$$

2.1.3 Log Returns

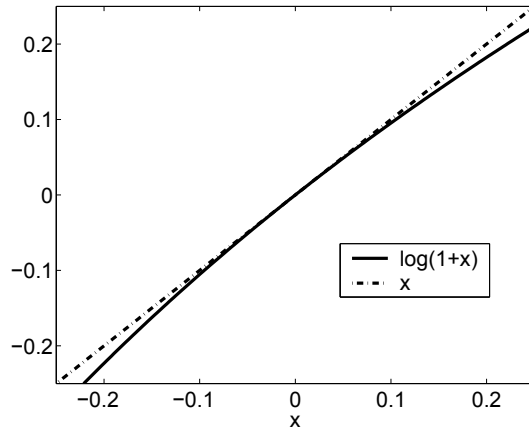


Fig. 2.1. Comparison of functions $\log(1+x)$ and x .

Log returns, also called *continuously compounded returns*, are denoted by r_t and defined as

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

where $p_t = \log(P_t)$ is called the *log price*.

Log returns are approximately equal to returns because if x is small, then $\log(1 + x) \approx x$, as can be seen in [Figure 2.1](#), where $\log(1 + x)$ is plotted. Notice in that figure that $\log(1 + x)$ is very close to x if $|x| < 0.1$, e.g., for returns that are less than 10%.

For example, a 5% return equals a 4.88% log return since $\log(1 + 0.05) = 0.0488$. Also, a -5% return equals a -5.13% log return since $\log(1 - 0.05) = -0.0513$. In both cases, $r_t = \log(1 + R_t) \approx R_t$. Also, $\log(1 + 0.01) = 0.00995$ and $\log(1 - 0.01) = -0.01005$, so log returns of $\pm 1\%$ are very close to the corresponding net returns.

One advantage of using log returns is simplicity of multiperiod returns. A k -period log return is simply the sum of the single-period log returns, rather than the product as for gross returns. To see this, note that the k -period log return is

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

2.1.4 Adjustment for Dividends

Many stocks, especially those of mature companies, pay dividends that must be accounted for when computing returns. Similarly, bonds pay interest. If a dividend (or interest) D_t is paid prior to time t , then the gross return at time t is defined as

$$1 + R_t = \frac{P_t + D_t}{P_{t-1}}, \quad (2.2)$$

and so the net return is $R_t = (P_t + D_t)/P_{t-1} - 1$ and the log return is $r_t = \log(1 + R_t) = \log(P_t + D_t) - \log(P_{t-1})$. Multiple-period gross returns are products of single-period gross returns so that

$$\begin{aligned} 1 + R_t(k) &= \left(\frac{P_t + D_t}{P_{t-1}}\right) \left(\frac{P_{t-1} + D_{t-1}}{P_{t-2}}\right) \cdots \left(\frac{P_{t-k+1} + D_{t-k+1}}{P_{t-k}}\right) \\ &= (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-k+1}), \end{aligned} \quad (2.3)$$

where, for any time s , $D_s = 0$ if there is no dividend between $s - 1$ and s . Similarly, a k -period log return is

$$\begin{aligned} r_t(k) &= \log\{1 + R_t(k)\} = \log(1 + R_t) + \cdots + \log(1 + R_{t-k+1}) \\ &= \log\left(\frac{P_t + D_t}{P_{t-1}}\right) + \cdots + \log\left(\frac{P_{t-k+1} + D_{t-k+1}}{P_{t-k}}\right). \end{aligned}$$

2.2 The Random Walk Model

The *random walk hypothesis* states that the single-period log returns, $r_t = \log(1 + R_t)$, are independent. Because

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

we have

$$\log\{1 + R_t(k)\} = r_t + \cdots + r_{t-k+1}. \quad (2.4)$$

It is sometimes assumed further that the log returns are $N(\mu, \sigma^2)$ for some constant mean and variance. Since sums of normal random variables are themselves normal, normality of single-period log returns implies normality of multiple-period log returns. Under these assumptions, $\log\{1 + R_t(k)\}$ is $N(k\mu, k\sigma^2)$.

2.2.1 Random Walks

Model (2.4) is an example of a random walk model. Let Z_1, Z_2, \dots be i.i.d. with mean μ and standard deviation σ . Let S_0 be an arbitrary starting point and

$$S_t = S_0 + Z_1 + \cdots + Z_t, \quad t \geq 1. \quad (2.5)$$

The process S_0, S_1, \dots is called a *random walk* and Z_1, Z_2, \dots are its steps. If the steps are normally distributed, then the process is called a *normal random walk*. The expectation and variance of S_t , conditional given S_0 , are $E(S_t|S_0) = S_0 + \mu t$ and $\text{Var}(S_t|S_0) = \sigma^2 t$. The parameter μ is called the *drift* and determines the general direction of the random walk. The parameter σ is the *volatility* and determines how much the random walk fluctuates about the conditional mean $S_0 + \mu t$. Since the standard deviation of S_t given S_0 is $\sigma\sqrt{t}$, $(S_0 + \mu t) \pm \sigma\sqrt{t}$ gives the mean plus and minus one standard deviation, which, for a normal random walk, gives a range containing 68% probability. The width of this range grows proportionally to \sqrt{t} , as is illustrated in [Figure 2.2](#), showing that at time $t = 0$ we know far less about where the random walk will be in the distant future compared to where it will be in the immediate future.

2.2.2 Geometric Random Walks

Recall that $\log\{1 + R_t(k)\} = r_t + \cdots + r_{t-k+1}$. Therefore,

$$\frac{P_t}{P_{t-k}} = 1 + R_t(k) = \exp(r_t + \cdots + r_{t-k+1}), \quad (2.6)$$

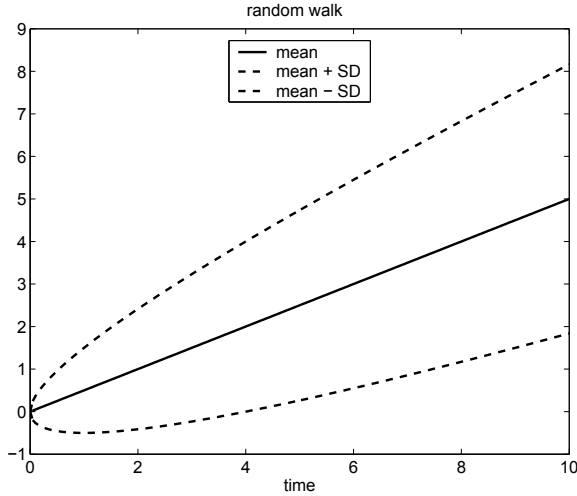


Fig. 2.2. Mean and bounds (mean plus and minus one standard deviation) on a random walk with $S_0 = 0$, $\mu = 0.5$, and $\sigma = 1$. At any given time, the probability of being between the bounds (dashed curves) is 68% if the distribution of the steps is normal.

so taking $k = t$, we have

$$P_t = P_0 \exp(r_t + r_{t-1} + \cdots + r_1). \quad (2.7)$$

We call such a process whose logarithm is a random walk a *geometric random walk* or an *exponential random walk*. If r_1, r_2, \dots are i.i.d. $N(\mu, \sigma^2)$, then P_t is lognormal for all t and the process is called a *lognormal geometric random walk with parameters* (μ, σ^2) .

2.2.3 Are Log Prices a Lognormal Geometric Random Walk?

Much work in mathematical finance assumes that prices follow a lognormal geometric random walk or its continuous-time analog, geometric Brownian motion. So a natural question is whether this assumption is usually true. The quick answer is “no.” The lognormal geometric random walk makes two assumptions: (1) the log returns are normally distributed and (2) the log returns are mutually independent.

In Chapters 4 and 5, we will investigate the marginal distributions of several series of log returns. The conclusion will be that, though the return density has a bell shape somewhat like that of normal densities, the tails of the log return distributions are generally much heavier than normal tails. Typically, a t -distribution with a small degrees-of-freedom parameter, say 4–6, is a much better fit than the normal model. However, the log-return distributions do appear to be symmetric, or at least nearly so.

The independence assumption is also violated. First, there is some correlation between returns. The correlations, however, are generally small. More seriously, returns exhibit *volatility clustering*, which means that if we see high volatility in current returns then we can expect this higher volatility to continue, at least for a while.

Before discarding the assumption that the prices of an asset are a lognormal geometric random walk, it is worth remembering that “all models are false, but some models are useful.” This assumption is sometimes useful, e.g., for deriving the famous Black–Scholes formula.

2.3 Bibliographic Notes

The random walk hypothesis is related to the so-called efficient market hypothesis; see Ruppert (2003) for discussion and further references. Bodie, Kane, and Marcus (1999) and Sharpe, Alexander, and Bailey (1995) are good introductions to the random walk hypothesis and market efficiency. A more advanced discussion of the random walk hypothesis is found in Chapter 2 of Campbell, Lo, and MacKinlay (1997) and Lo and MacKinlay (1999). Much empirical evidence about the behavior of returns is reviewed by Fama (1965, 1970, 1991, 1998). Evidence against the efficient market hypothesis can be found in the field of behavioral finance which uses the study of human behavior to understand market behavior; see Shefrin (2000), Shleifer (2000), and Thaler (1993). One indication of market inefficiency is excess volatility of market prices; see Shiller (1992) or Shiller (2000) for a less technical discussion.

Zuur, Ieno, Meesters, and Burg, D. (2009) is a good place to start learning R.

2.4 References

- Bodie, Z., Kane, A., and Marcus, A. (1999) *Investments*, 4th ed., Irwin/McGraw-Hill, Boston.
- Campbell, J., Lo, A., and MacKinlay, A. (1997) *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- Fama, E. (1965) The behavior of stock market prices. *Journal of Business*, **38**, 34–105.
- Fama, E. (1970) Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, 383–417.
- Fama, E. (1991) Efficient Capital Markets: II. *Journal of Finance*. **46**, 1575–1618.
- Fama, E. (1998) Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, **49**, 283–306.
- Lo, A. W., and MacKinlay, A. C. (1999) *A Non-Random Walk Down Wall Street*, Princeton University Press, Princeton and Oxford.

- Ruppert, D. (2003) *Statistics and Finance: An Introduction*, Springer, New York.
- Sharpe, W. F., Alexander, G. J., and Bailey, J. V. (1995) *Investments*, 6th ed., Simon and Schuster, Upper Saddle River, NJ.
- Shefrin, H. (2000) *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing*, Harvard Business School Press, Boston.
- Shiller, R. (1992) *Market Volatility*, Reprint ed., MIT Press, Cambridge, MA.
- Shiller, R. (2000) *Irrational Exuberance*, Broadway, New York.
- Shleifer, A. (2000) *Inefficient Markets: An Introduction to Behavioral Finance*, Oxford University Press, Oxford.
- Thaler, R. H. (1993) *Advances in Behavioral Finance*, Russell Sage Foundation, New York.
- Zuur, A., Ieno, E., Meesters, E., and Burg, D. (2009) *A Beginner's Guide to R*, Springer, New York.

2.5 R Lab

2.5.1 Data Analysis

Obtain the data set `Stock_FX_bond.csv` from the book's website and put it in your working directory. Start R and you should see a console window open up. Use **Change Dir** in the “File” menu to change to the working directory. Read the data with the following command:

```
dat = read.csv("Stock_bond.csv",header=TRUE)
```

The data set `Stock_FX_bond.csv` contains the volumes and adjusted closing (AC) prices of stocks and the S&P 500 (columns B–W), yields on bonds (columns X–AD).

This book does not give detailed information about R functions since this information is readily available elsewhere. For example, you can use R's help to obtain more information about the `read.csv` function by typing “`?read.csv`” in your R console and then hitting the Enter key. You should also use the manual *An Introduction to R* that is available on R's help file and also on CRAN. Another resource for those starting to learn R is Zuur et al. (2009).

An alternative to typing commands in the console is to start a new script from the “file” menu, put code into the editor, highlight the lines, and then type Ctrl-R to run the code that has been highlighted. This technique is useful for debugging. You can save the script file and then reuse or modify it.

Once a file is saved, the entire file can be run by “**sourcing**” it. You can use the “file” menu in R to source a file or use the `source` function. If the file is in the editor, then it can be run by hitting Ctrl-A to highlight the entire file and then Ctrl-R.

The next lines of code print the names of the variables in the data set, attach the data, and plot the adjusted closing prices of GM and Ford.

```
names(dat)
attach(dat)
par(mfrow=c(1,2))
plot(GM_AC)
plot(F_AC)
```

The R function `attach` puts a database into the R search path. This means that the database is searched by R when evaluating a variable, so objects in the database can be accessed by simply giving their names. The function `par` specifies plotting parameters and `mfrow=c(n1,n2)` specifies “make a figure, fill by rows, n1 rows and n2 columns.” Thus, the first n1 plots fill the first row and so forth. `mfc col(n1,n2)` fills by columns and so would put the first n2 plots in the first column. As mentioned before, more information about these and other R functions can be obtained from R’s online help or the manual *An Introduction to R*.

Run the code below to find the sample size (`n`), compute GM and Ford returns, and plot GM returns versus the Ford returns.

```
n = dim(dat)[1]
GMReturn = GM_AC[2:n]/GM_AC[1:(n-1)] - 1
FReturn = F_AC[2:n]/F_AC[1:(n-1)] - 1
par(mfrow=c(1,1))
plot(GMReturn,FReturn)
```

Problem 1 *Do the GM and Ford returns seem positively correlated? Do you notice any outlying returns? If “yes,” do outlying GM returns seem to occur with outlying Ford returns?*

Problem 2 *Compute the log returns for GM and plot the returns versus the log returns? How highly correlated are the two types of returns? (The R function `cor` computes correlations.)*

When you exit R, you can “Save workspace image,” which will create an R workspace file in your working directory. Later, you can restart R from within Windows™ and load this workspace image into memory by right-clicking on the R workspace file. When R starts, your working directory will be the folder containing the R workspace that was opened.

2.5.2 Simulations

Hedge funds can earn high profits by the use of leverage, but leverage also creates high risk. The simulations in this section explore the effects of leverage.

Suppose a hedge fund owns \$1,000,000 of stock and used \$50,000 of its own capital and \$950,000 in borrowed money for the purchase. If the value of the stock falls below \$950,000 at the end of any trading day, then the hedge

fund must sell all the stock and repay the loan. This will wipe out its \$50,000 investment. The hedge fund is said to be leveraged 20:1 since its position is 20 times the amount of its own capital invested.

The daily log returns on the stock have a mean of 0.05/year and a standard deviation of 0.23/year. These can be converted to rates per trading day by dividing by 253 and $\sqrt{253}$, respectively.

Problem 3 *What is the probability that the value of the stock will be below \$950,000 at the close of at least one of the next 45 trading days? To answer this question, run the code below.*

```
niter = 1e5          # number of iterations
below = rep(0,niter) # set up storage
set.seed(2009)
for (i in 1:niter)
{
  r = rnorm(45,mean=.05/253,
           sd=.23/sqrt(253)) # generate random numbers
  logPrice = log(1e6) + cumsum(r)
  minlogP = min(logPrice) # minimum price over next 45 days
  below[i] = as.numeric(minlogP < log(950000))
}
mean(below)
```

If you are unfamiliar with any of the R functions used here, then use R's help to learn about them; e.g., type `?rnorm` to learn that `rnorm` generates normally distributed random numbers. You should study each line of code, understand what it is doing, and convince yourself that the code estimates the probability being requested. Note that anything that follows a pound sign is a comment and is used only to annotate the code.

Suppose the hedge fund will sell the stock for a profit of at least \$100,000 if the value of the stock rises to at least \$1,100,000 at the end of one of the first 100 trading days, sell it for a loss if the value falls below \$950,000 at the end of one of the first 100 trading days, or sell after 100 trading days if the closing price has stayed between \$950,000 and \$1,000,000.

The following questions can be answered by simulations much like the one above. Ignore trading costs and interest when answering these questions.

Problem 4 *What is the probability that the hedge fund will make a profit of at least \$100,000?*

Problem 5 *What is the probability the hedge fund will suffer a loss?*

Problem 6 *What is the expected profit from this trading strategy?*

Problem 7 *What is the expected return? When answering this question, remember that only \$50,000 was invested. Also, the units of return are time, e.g., one can express a return as a daily return or a weekly return. Therefore, one must keep track of how long the hedge fund holds its position before selling.*

2.6 Exercises

- The daily log returns on a stock are independent and normally distributed with mean 0.001 and standard deviation 0.015. Suppose you buy \$1000 worth of this stock.
 - What is the probability that after one trading day your investment is worth less than \$990? (**Note:** The R function `pnorm` will compute a normal CDF, so, for example, `pnorm(0.3,mean=0.1,sd=0.2)` is the normal CDF with mean 0.1 and standard deviation 0.2 evaluated at 0.3.)
 - What is the probability that after five trading days your investment is worth less than \$990?
- The yearly log returns on a stock are normally distributed with mean 0.1 and standard deviation 0.2. The stock is selling at \$100 today. What is the probability that one year from now it is selling at \$110 or more?
- Suppose the price of a stock at times 1, 2, and 3 are $P_1 = 95$, $P_2 = 103$, and $P_3 = 98$. Find $r_3(2)$.
- The prices and dividends of a stock are given in the table below.
 - What is R_2 ?
 - What is $R_4(3)$?
 - What is r_3 ?

| t | P_t | D_t |
|-----|-------|-------|
| 1 | 52 | 0.2 |
| 2 | 54 | 0.2 |
| 3 | 53 | 0.2 |
| 4 | 59 | 0.25 |

- Let r_t be a log return. Suppose that r_1, r_2, \dots are i.i.d. $N(0.06, 0.47)$.
 - What is the distribution of $r_t(4) = r_t + r_{t-1} + r_{t-2} + r_{t-3}$?
 - What is $P\{r_1(4) < 2\}$?
 - What is the covariance between $r_1(2)$ and $r_2(2)$?
 - What is the conditional distribution of $r_t(3)$ given $r_{t-2} = 0.6$?
- Suppose that X_1, X_2, \dots is a lognormal geometric random walk with parameters (μ, σ^2) . More specifically, suppose that $X_k = X_0 \exp(r_1 + \dots + r_k)$, where X_0 is a fixed constant and r_1, r_2, \dots are i.i.d. $N(\mu, \sigma^2)$.
 - Find $P(X_2 > 1.3 X_0)$.
 - Use (A.4) to find the density of X_1 .
 - Find a formula for the 0.9 quantile of X_k for all k .

- (d) What is the expected value of X_k^2 for any k ? (Find a formula giving the expected value as a function of k .)
 - (e) Find the variance of X_k for any k .
7. The daily log returns on a stock are normally distributed with mean 0.0002 and standard deviation 0.03. The stock price is now \$97. What is the probability that it will exceed \$100 after 20 trading days?



<http://www.springer.com/978-1-4419-7787-8>

Statistics and Data Analysis for Financial Engineering

Ruppert, D.

2011, XXII, 638 p.,

ISBN: 978-1-4419-7787-8