2

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 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 $P_t \geq 0$,

D. Ruppert, D.S. Matteson, *Statistics and Data Analysis for Financial Engineering*, Springer Texts in Statistics,
DOI 10.1007/978-1-4939-2614-5_2

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

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 %. One's final wealth at time t is one's initial wealth at time $t-1$ times the gross return. Stated differently, if X_0 is the initial at time $t-1$, then $X_0(1 + R_t)$ is one's wealth at time t.

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 this example, if t is measured in years, then, stated more precisely, the 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):

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

The k-period net return is $R_t(k)$.

2.1.3 Log Returns

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 been seen in Fig. 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 ± 1 % are very close to the corresponding net returns. Since returns are smaller in magnitude over shorter periods, we can expect returns and log returns to be similar for daily returns, less similar for yearly returns, and not necessarily similar for longer periods such as 10 years.

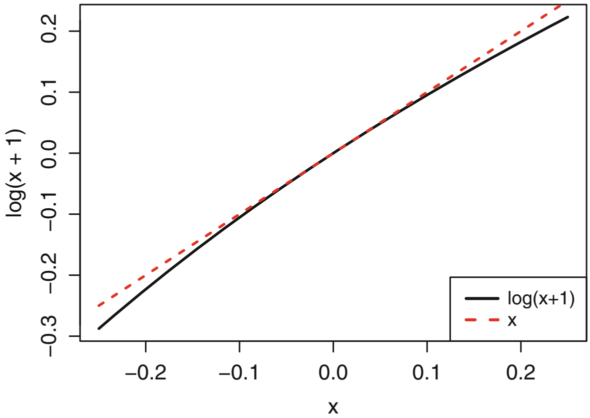

Fig. 2.1. *Comparison of functions* $\log(1+x)$ *and* x*.*

The return and log return have the same sign. The magnitude of the log return is smaller (larger) than that of the return if they are both positive (negative). The difference between a return and a log return is most pronounced when both are very negative. Returns close to the lower bound of -1, that is complete losses, correspond to log return close to $-\infty$.

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

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

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

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

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

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

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 $Z_1, Z_2, \ldots$ be i.i.d. (independent and identically distributed) with mean μ and standard deviation σ. Let S_0 be an arbitrary starting point and

$$S_t = S_0 + Z_1 + \cdots + Z_t, \quad t \geq 1. \tag{2.5}$$

From (2.5), S_t is the position of the random walker after t steps starting at S_0.

The process $S_0, S_1, \ldots$ is called a *random walk* and $Z_1, Z_2, \ldots$ 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 $\mathrm{Var}(S_t|S_0) = \sigma^2 t$. 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 Fig. 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,

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

so taking $k = t$, we have

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

We call such a process whose logarithm is a random walk a *geometric random walk* or an *exponential random walk.* If $r_1, r_2, \ldots$ 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). As discussed in Appendix A.9.4, μ is called the log-mean and σ is called the log-standard deviation of the log-normal distribution of $\exp(r_t)$. Also, μ is sometimes called the log-drift of the lognormal geometric random walk.

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 Chaps. 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

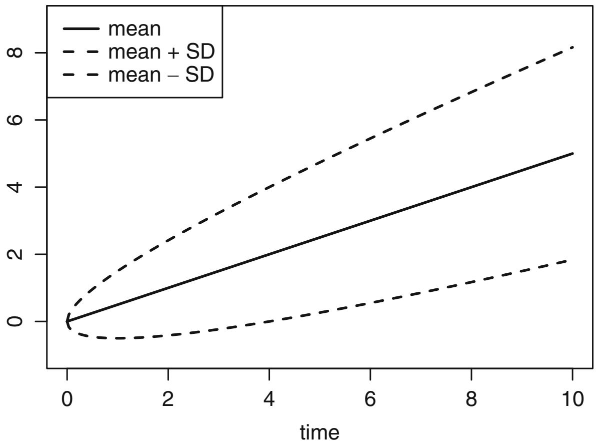

Fig. 2.2. *Mean and bounds (mean plus and minus one standard deviation) on a random walk with* $S_0 = 0$*,* $\mu = 0.5$*, and* $\sigma = 1$*. At any given time, the probability of being between the bounds (dashed curves) is* 68 % *if the distribution of the steps is normal. Since* $\mu > 0$*, there is an overall positive trend that would be reversed if* μ *were negative.*

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. Volatility clustering can be detected by checking for correlations between the *squared* returns.

Before discarding the assumption that the prices of an asset are a lognormal geometric random walk, it is worth remembering Box's dictum 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 et al. (2003) for discussion and further references. Bodie et al. (1999) and Sharpe et al. (1995) are good introductions to the random walk hypothesis and market efficiency. A more advanced discussion of the random walk hypothesis is found in Chap. 2 of Campbell et al. (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.

R will be used extensively in what follows. Dalgaard (2008) and Zuur et al. (2009) are good places to start learning R.

2.4 R Lab

2.4.1 Data Analysis

Obtain the data set `Stock_bond.csv` from the book's website and put it in your working directory. Start R[1] 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_bond.csv` contains daily volumes and adjusted closing (AC) prices of stocks and the S&P 500 (columns B–W) and yields on bonds (columns X–AD) from 2-Jan-1987 to 1-Sep-2006.

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 press Ctrl-R to run the code that has been highlighted.[2] 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.

[1] You can also run R from `Rstudio` and, in fact, `Rstudio` is highly recommended. The authors switched from R to `Rstudio` while the second edition of this book was being written.

[2] Or click the "run" button in Rstudio.

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

Here and elsewhere in this book, line numbers are often added when listing `R` code. The line numbers are not part of the code.

By default, as in lines 4 and 5, points are plotted with the character “o”. To plot a line instead, use, for example `plot(GM_AC, type = "l")`. Similarly, `plot(GM_AC, type = "b")` plots both points and a line.

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. If `dat` was not attached, then line 4 would be replaced by `plot(dat$GM_AC)` and similarly for line 5.

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. `mfcol(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 net returns versus the Ford returns.

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

On lines 2 and 3, the index `-1` means all indices except the first and similarly `-n` means all indices except the last.

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.)*

Problem 3 *Repeat Problem 1 with Microsoft (MSFT) and Merck (MRK).*

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 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. A useful trick when starting a project in a new folder is to put an empty saved workspace into this folder. Double-clicking on the workspace starts R with the folder as the working directory.

2.4.2 Simulations

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

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. Suppose that if the value of the stock falls below \$950,000 at the end of any trading day, then the hedge fund will 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.

Suppose that 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 4 *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.*

```
1  niter = 1e5           #  number of iterations
2  below = rep(0, niter)  # set up storage
3  set.seed(2009)
4  for (i in 1:niter)
5  {
6     r = rnorm(45, mean = 0.05/253,
7        sd = 0.23/sqrt(253)) # generate random numbers
8     logPrice = log(1e6) + cumsum(r)
9      minlogP = min(logPrice) # minimum price over next 45 days
10    below[i] = as.numeric(minlogP < log(950000))
11 }
12 mean(below)
```

On line 10, `below[i]` equals 1 if, for the `i`th simulation, the minimum price over 45 days is less that 950,000. Therefore, on line 12, `mean(below)` is the proportion of simulations where the minimum price is less than 950,000.

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 $1,100,000.

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

Problem 5 *What is the probability that the hedge fund will make a profit of at least $100,000?*

Problem 6 *What is the probability the hedge fund will suffer a loss?*

Problem 7 *What is the expected profit from this trading strategy?*

Problem 8 *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.*

2.4.3 Simulating a Geometric Random Walk

In this section you will use simulations to see how stock prices evolve when the log-returns are i.i.d. normal, which implies that the price series is a geometric random walk.

Run the following `R` code. The `set.seed()` command insures that everyone using this code will have the same random numbers and will obtain the same price series. There are 253 trading days per year, so you are simulating 1 year of daily returns nine times. The price starts at 120.

The code `par(mfrow=c(3,3))` on line 3 opens a graphics window with three rows and three columns and `rnorm()` on line 6 generates normally distributed random numbers.

```
1 set.seed(2012)
2 n = 253
3 par(mfrow=c(3,3))
4 for (i in (1:9))
5 {
6    logr = rnorm(n, 0.05 / 253, 0.2 / sqrt(253))
```

```
    price = c(120, 120 * exp(cumsum(logr)))
    plot(price, type = "b")
}
```

Problem 9 *In this simulation, what are the mean and standard deviation of the log-returns for 1 year?*

Problem 10 *Discuss how the price series appear to have momentum. Is the appearance of momentum real or an illusion?*

Problem 11 *Explain what the code* `c(120,120*exp(cumsum(logr)))` *does.*

2.4.4 Let's Look at McDonald's Stock

In this section we will be looking at daily returns on McDonald's stock over the period 2010–2014. To start the lab, run the following commands to get daily adjusted prices over this period:

```
data = read.csv('MCD_PriceDaily.csv')
head(data)
adjPrice = data[, 7]
```

Problem 12 *Compute the returns and log returns and plot them against each other. As discussed in Sect. 2.1.3, does it seem reasonable that the two types of daily returns are approximately equal?*

Problem 13 *Compute the mean and standard deviation for both the returns and the log returns. Comment on the similarities and differences you perceive in the first two moments of each random variable. Does it seem reasonable that they are the same?*

Problem 14 *Perform a t-test to compare the means of the returns and the log returns. Comment on your findings. Do you reject the null hypothesis that they are the same mean at 5 % significance? Or do you accept it? [Hint: Should you be using an independent samples t-test or a paired-samples t-test?]*
What are the assumptions behind the t-test? Do you think that they are met in this example? If the assumptions made by the t-test are not met, how would this affect your interpretation of the results of the test?

Problem 15 *After looking at return and log return data for McDonald's, are you satisfied that for small values, log returns and returns are interchangeable?*

Problem 16 *Assume that McDonald's log returns are normally distributed with mean and standard deviation equal to their estimates and that you have been made the following proposition by a friend: If at any point within the next 20 trading days, the price of McDonald's falls below 85 dollars, you will be paid \$100, but if it does not, you have to pay him \$1. The current price of McDonald's is at the end of the sample data, \$93.07. Are you willing to make the bet? (Use 10,000 iterations in your simulation and use the command* `set.seed(2015)` *to ensure your results are the same as the answer key)*

Problem 17 *After coming back to your friend with an unwillingness to make the bet, he asks you if you are willing to try a slightly different deal. This time the offer stays the same as before, except he would pay an additional \$25 if the price ever fell below \$84.50. You still only pay him \$1 for losing. Do you now make the bet?*

2.5 Exercises

1. Suppose that the daily log returns on a stock are independent and normally distributed with mean 0.001 and standard deviation 0.015. Suppose you buy \$1,000 worth of this stock.
   (a) 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.)
   (b) What is the probability that after five trading days your investment is worth less than \$990?
2. 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 1 year from now it is selling at \$110 or more?
3. The yearly log returns on a stock are normally distributed with mean 0.08 and standard deviation 0.15. The stock is selling at \$80 today. What is the probability that 2 years from now it is selling at \$90 or more?
4. Suppose the prices of a stock at times 1, 2, and 3 are $P_1 = 95$, $P_2 = 103$, and $P_3 = 98$. Find $r_3(2)$.
5. The prices and dividends of a stock are given in the table below.
   (a) What is R_2?
   (b) What is $R_4(3)$?
   (c) 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

6. The prices and dividends of a stock are given in the table below.
   (a) Find $R_3(2)$,
   (b) Find $r_4(3)$.

t	P_t	D_t
1	82	0.1
2	85	0.1
3	83	0.1
4	87	0.125

7. Let r_t be a log return. Suppose that $r_1, r_2, \ldots$ are i.i.d. $N(0.06, 0.47)$.
   (a) What is the distribution of $r_t(4) = r_t + r_{t-1} + r_{t-2} + r_{t-3}$?
   (b) What is $P\{r_1(4) < 2\}$?
   (c) What is the covariance between $r_2(1)$ and $r_2(2)$?
   (d) What is the conditional distribution of $r_t(3)$ given $r_{t-2} = 0.6$?
8. Suppose that $X_1, X_2, \ldots$ is a lognormal geometric random walk with parameters (μ, σ^2). More specifically, suppose that $X_k = X_0 \exp(r_1 + \cdots + r_k)$, where X_0 is a fixed constant and $r_1, r_2, \ldots$ are i.i.d. $N(\mu, \sigma^2)$.
   (a) Find $P(X_2 > 1.3\, X_0)$.
   (b) Use (A.4) to find the density of X_1.
   (c) 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.
9. Suppose that $X_1, X_2, \ldots$ is a lognormal geometric random walk with parameters $\mu = 0.1, \sigma = 0.2$.
   (a) Find $P(X_3 > 1.2X_0)$.
   (b) Find the conditional variance of X_k/k given X_0 for any k.
   (c) Find the minimum number of days before the probability is at least 0.9 of doubling one's money, that is, find the small value of t such that $P(P_t/P_0 \geq 2) \geq 0.9$.
10. 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?
11. Suppose that daily log-returns are $N(0.0005, 0.012)$. Find the smallest value of t such that $P(P_t/P_0 \geq 2) \geq 0.9$, that is, that after t days the probability the price has doubled is at least 90 %.

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.

Dalgaard, P. (2008) *Introductory Statistics with* R*, 2nd ed.*, Springer.

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.

http://www.springer.com/978-1-4939-2613-8

Statistics and Data Analysis for Financial Engineering
with R examples
Ruppert, D.; Matteson, D.S.
2015, XXVI, 719 p. 221 illus., 108 illus. in color.,
Hardcover
ISBN: 978-1-4939-2613-8