Introduction to Computational Finance and Financial Econometrics Chapter 1 Asset Return Calculations

Eric Zivot
Department of Economics, University of Washington

December 31, 1998
Updated: January 7, 2002

1 The Time Value of Money

Consider an amount \$V invested for n years at a *simple interest rate* of R per annum (where R is expressed as a decimal). If compounding takes place only at the end of the year the *future value* after n years is

$$FV_n = \$V \cdot (1 + R)^n.$$

Example 1 *Consider putting \$1000 in an interest checking account that pays a simple annual percentage rate of* 3%. *The future value after* $n = 1, 5$ *and* 10 *years is, respectively,*

$$\begin{aligned} FV_1 &= \$1000 \cdot (1.03)^1 = \$1030 \\ FV_5 &= \$1000 \cdot (1.03)^5 = \$1159.27 \\ FV_{10} &= \$1000 \cdot (1.03)^{10} = \$1343.92. \end{aligned}$$

If interest is paid m time per year then the future value after n years is

$$FV_n^m = \$V \cdot \left(1 + \frac{R}{m}\right)^{m \cdot n}.$$

$\frac{R}{m}$ is often referred to as the *periodic interest rate*. As m, the frequency of compounding, increases the rate becomes continuously compounded and it can be shown that future value becomes

$$FV_n^c = \lim_{m \to \infty} \$V \cdot \left(1 + \frac{R}{m}\right)^{m \cdot n} = \$V \cdot e^{R \cdot n},$$

where $e^{(\cdot)}$ is the exponential function and $e^1 = 2.71828$.

Example 2 *If the simple annual percentage rate is 10% then the value of \$1000 at the end of one year ($n = 1$) for different values of m is given in the table below.*

Compounding Frequency	Value of \$1000 at end of 1 year ($R = 10\%$)
Annually ($m = 1$)	1100
Quarterly ($m = 4$)	1103.8
Weekly ($m = 52$)	1105.1
Daily ($m = 365$)	1105.515
Continuously ($m = \infty$)	1105.517

We now consider the relationship between simple interest rates, periodic rates, effective annual rates and continuously compounded rates. Suppose an investment pays a periodic interest rate of 2% each quarter. This gives rise to a simple annual rate of 8% (2% $\times 4$ quarters). At the end of the year, \$1000 invested accrues to

$$\$1000 \cdot \left(1 + \frac{0.08}{4}\right)^{4 \cdot 1} = \$1082.40.$$

The *effective annual rate*, R_A, on the investment is determined by the relationship

$$\$1000 \cdot (1 + R_A) = \$1082.40,$$

which gives $R_A = 8.24\%$. The effective annual rate is greater than the simple annual rate due to the payment of interest on interest.

The general relationship between the simple annual rate R with payments m time per year and the effective annual rate, R_A, is

$$(1 + R_A) = \left(1 + \frac{R}{m}\right)^{m \cdot 1}.$$

Example 3 *To determine the simple annual rate with quarterly payments that produces an effective annual rate of* 12%, *we solve*

$$
\begin{aligned}
1.12 &= \left(1+\frac{R}{4}\right)^4 \Longrightarrow \\
R &= \left((1.12)^{1/4}-1\right)\cdot 4 \\
&= 0.0287\cdot 4 \\
&= 0.1148
\end{aligned}
$$

Suppose we wish to calculate a value for a continuously compounded rate, R_c, when we know the $m-$period simple rate R. The relationship between such rates is given by

$$e^{R_c} = \left(1+\frac{R}{m}\right)^m. \tag{1}$$

Solving (1) for R_c gives

$$R_c = m\ln\left(1+\frac{R}{m}\right), \tag{2}$$

and solving (1) for R gives

$$R = m\left(e^{R_c/m}-1\right). \tag{3}$$

Example 4 *Suppose an investment pays a periodic interest rate of 5% every six months ($m=2, R/2=0.05$). In the market this would be quoted as having an annual percentage rate of 10%. An investment of \$100 yields* $\$100\cdot(1.05)^2 = \$110.25$ *after one year. The effective annual rate is then* 10.25%. *Suppose we wish to convert the simple annual rate of $R=10\%$ to an equivalent continuously compounded rate. Using (2) with $m=2$ gives*

$$R_c = 2\cdot\ln(1.05) = 0.09758.$$

That is, if interest is compounded continuously at an annual rate of 9.758% *then \$100 invested today would grow to* $\$100\cdot e^{0.09758} = \$110.25$.

2 Asset Return Calculations

2.1 Simple Returns

Let P_t denote the price in month t of an asset that pays no dividends and let P_{t-1} denote the price in month $t-1$[1]. Then the one month *simple net return* on an investment in the asset between months $t-1$ and t is defined as

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \%\Delta P_t. \tag{4}$$

Writing $\frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$, we can define the *simple gross return* as

$$1 + R_t = \frac{P_t}{P_{t-1}}. \tag{5}$$

Notice that the one month gross return has the interpretation of the future value of \$1 invested in the asset for one month. Unless otherwise stated, when we refer to returns we mean net returns.

(mention that simple returns cannot be less than 1 (100%) since prices cannot be negative)

Example 5 *Consider a one month investment in Microsoft stock. Suppose you buy the stock in month* $t-1$ *at* $P_{t-1} = \$85$ *and sell the stock the next month for* $P_t = \$90$. *Further assume that Microsoft does not pay a dividend between months* $t-1$ *and* t. *The one month simple net and gross returns are then*

$$\begin{aligned} R_t &= \frac{\$90 - \$85}{\$85} = \frac{\$90}{\$85} - 1 = 1.0588 - 1 = 0.0588, \\ 1 + R_t &= 1.0588. \end{aligned}$$

The one month investment in Microsoft yielded a 5.88% *per month return. Alternatively, \$1 invested in Microsoft stock in month* $t-1$ *grew to* \$1.0588 *in month* t.

2.2 Multi-period returns

The simple two-month return on an investment in an asset between months $t-2$ and t is defined as

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

[1] We make the convention that the default investment horizon is one month and that the price is the closing price at the end of the month. This is completely arbitrary and is used only to simplify calculations.

Since $\frac{P_t}{P_{t-2}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}}$ the two-month return can be rewritten as

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

Then the simple two-month gross return becomes

$$1 + R_t(2) = (1+R_t)(1+R_{t-1}) = 1 + R_{t-1} + R_t + R_{t-1}R_t,$$

which is a *geometric* (multiplicative) sum of the two simple one-month gross returns and not the simple sum of the one month returns. If, however, R_{t-1} and R_t are small then $R_{t-1}R_t \approx 0$ and $1 + R_t(2) \approx 1 + R_{t-1} + R_t$ so that $R_t(2) \approx R_{t-1} + R_t$.

In general, the k-month gross return is defined as the geometric average of k one month gross returns

$$\begin{aligned} 1 + R_t(k) &= (1+R_t)(1+R_{t-1}) \cdots (1+R_{t-k+1}) \\ &= \prod_{j=0}^{k-1} (1 + R_{t-j}). \end{aligned}$$

Example 6 *Continuing with the previous example, suppose that the price of Microsoft stock in month $t-2$ is \$80 and no dividend is paid between months $t-2$ and t. The two month net return is*

$$R_t(2) = \frac{\$90 - \$80}{\$80} = \frac{\$90}{\$80} - 1 = 1.1250 - 1 = 0.1250,$$

or 12.50% per two months. The two one month returns are

$$\begin{aligned} R_{t-1} &= \frac{\$85 - \$80}{\$80} = 1.0625 - 1 = 0.0625 \\ R_t &= \frac{\$90 - 85}{\$85} = 1.0588 - 1 = 0.0588, \end{aligned}$$

and the geometric average of the two one month gross returns is

$$1 + R_t(2) = 1.0625 \times 1.0588 = 1.1250.$$

2.3 Annualizing returns

Very often returns over different horizons are annualized, i.e. converted to an annual return, to facilitate comparisons with other investments. The annualization process depends on the holding period of the investment and an implicit assumption about compounding. We illustrate with several examples.

To start, if our investment horizon is one year then the annual gross and net returns are just

$$
\begin{aligned}
1 + R_A &= 1 + R_t(12) = \frac{P_t}{P_{t-12}} = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-11}), \\
R_A &= \frac{P_t}{P_{t-12}} - 1 = (1 + R_t)(1 + R_{t-1}) \cdots (1 + R_{t-11}) - 1.
\end{aligned}
$$

In this case, no compounding is required to create an annual return.

Next, consider a one month investment in an asset with return R_t. What is the annualized return on this investment? If we assume that we receive the same return $R = R_t$ every month for the year then the gross 12 month or gross annual return is

$$1 + R_A = 1 + R_t(12) = (1 + R)^{12}.$$

Notice that the annual gross return is defined as the monthly return compounded for 12 months. The net annual return is then

$$R_A = (1 + R)^{12} - 1.$$

Example 7 *In the first example, the one month return,* R_t*, on Microsoft stock was* 5.88%*. If we assume that we can get this return for 12 months then the annualized return is*

$$R_A = (1.0588)^{12} - 1 = 1.9850 - 1 = 0.9850$$

or 98.50% *per year. Pretty good!*

Now, consider a two month investment with return $R_t(2)$. If we assume that we receive the same two month return $R(2) = R_t(2)$ for the next 6 two month periods then the gross and net annual returns are

$$
\begin{aligned}
1 + R_A &= (1 + R(2))^6, \\
R_A &= (1 + R(2))^6 - 1.
\end{aligned}
$$

Here the annual gross return is defined as the two month return compounded for 6 months.

Example 8 *In the second example, the two month return,* $R_t(2)$*, on Microsoft stock was* 12.5%*. If we assume that we can get this two month return for the next 6 two month periods then the annualized return is*

$$R_A = (1.1250)^6 - 1 = 2.0273 - 1 = 1.0273$$

or 102.73% per year.

To complicate matters, now suppose that our investment horizon is two years. That is we start our investment at time $t-24$ and cash out at time t. The two year gross return is then $1 + R_t(24) = \frac{P_t}{P_{t-24}}$. What is the annual return on this two year investment? To determine the annual return we solve the following relationship for R_A :

$$\begin{aligned} (1 + R_A)^2 &= 1 + R_t(24) \Longrightarrow \\ R_A &= (1 + R_t(24))^{1/2} - 1. \end{aligned}$$

In this case, the annual return is compounded twice to get the two year return and the relationship is then solved for the annual return.

Example 9 *Suppose that the price of Microsoft stock* 24 *months ago is* $P_{t-24} = \$50$ *and the price today is* $P_t = \$90$*. The two year gross return is* $1+R_t(24) = \frac{\$90}{\$50} = 1.8000$ *which yields a two year net return of* $R_t(24) = 80\%$*. The annual return for this investment is defined as*

$$R_A = (1.800)^{1/2} - 1 = 1.3416 - 1 = 0.3416$$

or 34.16% per year.

2.4 Adjusting for dividends

If an asset pays a dividend, D_t, sometime between months $t-1$ and t, the return calculation becomes

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

where $\frac{P_t - P_{t-1}}{P_{t-1}}$ is referred as the *capital gain* and $\frac{D_t}{P_{t-1}}$ is referred to as the *dividend yield.*

3 Continuously Compounded Returns

3.1 One Period Returns

Let R_t denote the simple monthly return on an investment. The *continuously compounded monthly return,* r_t, is defined as

$$r_t = \ln(1 + R_t) = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{6}$$

where $\ln(\cdot)$ is the natural log function[2]. To see why r_t is called the continuously compounded return, take the exponential of both sides of (6) to give

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

Rearranging we get

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

so that r_t is the continuously compounded growth rate in prices between months $t-1$ and t. This is to be contrasted with R_t which is the simple growth rate in prices between months $t-1$ and t without any compounding. Furthermore, since $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$ it follows that

$$\begin{aligned} r_t &= \ln\left(\frac{P_t}{P_{t-1}}\right) \\ &= \ln(P_t) - \ln(P_{t-1}) \\ &= p_t - p_{t-1} \end{aligned}$$

where $p_t = \ln(P_t)$. Hence, the continuously compounded monthly return, r_t, can be computed simply by taking the first difference of the natural logarithms of monthly prices.

Example 10 *Using the price and return data from example 1, the continuously compounded monthly return on Microsoft stock can be computed in two ways:*

$$r_t = \ln(1.0588) = 0.0571$$

[2] The continuously compounded return is always defined since asset prices, P_t, are always non-negative. Properties of logarithms and exponentials are discussed in the appendix to this chapter.

or

$$r_t = \ln(90) - \ln(85) = 4.4998 - 4.4427 = 0.0571.$$

Notice that r_t is slightly smaller than R_t. Why?

Given a monthly continuously compounded return r_t, is straightforward to solve back for the corresponding simple net return R_t :

$$R_t = e^{r_t} - 1$$

Hence, nothing is lost by considering continuously compounded returns instead of simple returns.

Example 11 *In the previous example, the continuously compounded monthly return on Microsoft stock is $r_t = 5.71\%$. The implied simple net return is then*

$$R_t = e^{.0571} - 1 = 0.0588.$$

Continuously compounded returns are very similar to simple returns as long as the return is relatively small, which it generally will be for monthly or daily returns. For modeling and statistical purposes, however, it is much more convenient to use continuously compounded returns due to the additivity property of multiperiod continuously compounded returns and unless noted otherwise from here on we will work with continuously compounded returns.

3.2 Multi-Period Returns

The computation of multi-period continuously compounded returns is considerably easier than the computation of multi-period simple returns. To illustrate, consider the two month continuously compounded return defined as

$$r_t(2) = \ln(1 + R_t(2)) = \ln\left(\frac{P_t}{P_{t-2}}\right) = p_t - p_{t-2}.$$

Taking exponentials of both sides shows that

$$P_t = P_{t-2}e^{r_t(2)}$$

so that $r_t(2)$ is the continuously compounded growth rate of prices between months $t-2$ and t. Using $\frac{P_t}{P_{t-2}} = \frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}}$ and the fact that $\ln(x \cdot y) = \ln(x) + \ln(y)$ it follows that

$$\begin{aligned} r_t(2) &= \ln\left(\frac{P_t}{P_{t-1}} \cdot \frac{P_{t-1}}{P_{t-2}}\right) \\ &= \ln\left(\frac{P_t}{P_{t-1}}\right) + \ln\left(\frac{P_{t-1}}{P_{t-2}}\right) \\ &= r_t + r_{t-1}. \end{aligned}$$

Hence the continuously compounded two month return is just the sum of the two continuously compounded one month returns. Recall that with simple returns the two month return is of a multiplicative form (geometric average).

Example 12 *Using the data from example 2, the continuously compounded two month return on Microsoft stock can be computed in two equivalent ways. The first way uses the difference in the logs of* P_t *and* P_{t-2}*:*

$$r_t(2) = \ln(90) - \ln(80) = 4.4998 - 4.3820 = 0.1178.$$

The second way uses the sum of the two continuously compounded one month returns. Here $r_t = \ln(90) - \ln(85) = 0.0571$ *and* $r_{t-1} = \ln(85) - \ln(80) = 0.0607$ *so that*

$$r_t(2) = 0.0571 + 0.0607 = 0.1178.$$

Notice that $r_t(2) = 0.1178 < R_t(2) = 0.1250.$

The continuously compounded $k-$month return is defined by

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

Using similar manipulations to the ones used for the continuously compounded two month return we may express the continuously compounded $k-$month return as the sum of k continuously compounded monthly returns:

$$r_t(k) = \sum_{j=0}^{k-1} r_{t-j}.$$

The additivitity of continuously compounded returns to form multiperiod returns is an important property for statistical modeling purposes.

3.3 Annualizing Continuously Compounded Returns

Just as we annualized simple monthly returns, we can also annualize continuously compounded monthly returns.

To start, if our investment horizon is one year then the annual continuously compounded return is simply the sum of the twelve monthly continuously compounded returns

$$\begin{aligned} r_A &= r_t(12) = r_t + r_{t-1} + \cdots + r_{t-11} \\ &= \sum_{j=0}^{11} r_{t-j}. \end{aligned}$$

Define the average continuously compounded monthly return to be

$$\overline{r}_m = \frac{1}{12} \sum_{j=0}^{11} r_{t-j}.$$

Notice that

$$12 \cdot \overline{r}_m = \sum_{j=0}^{11} r_{t-j}$$

so that we may alternatively express r_A as

$$r_A = 12 \cdot \overline{r}_m.$$

That is, the continuously compounded annual return is 12 times the average of the continuously compounded monthly returns.

Next, consider a one month investment in an asset with continuously compounded return r_t. What is the continuously compounded annual return on this investment? If we assume that we receive the same return $r = r_t$ every month for the year then $r_A = r_t(12) = 12 \cdot r$.

4 Further Reading

This chapter describes basic asset return calculations with an emphasis on equity calculations. Campbell, Lo and MacKinlay provide a nice treatment of continuously compounded returns. A useful summary of a broad range of return calculations is given in Watsham and Parramore (1998). A comprehensive treatment of fixed income return calculations is given in Stigum (1981) and the official source of fixed income calculations is “The Pink Book”.

5 Appendix: Properties of exponentials and logarithms

The computation of continuously compounded returns requires the use of natural logarithms. The natural logarithm function, $\ln(\cdot)$, is the inverse of the exponential function, $e^{(\cdot)} = \exp(\cdot)$, where $e^1 = 2.718$. That is, $\ln(x)$ is defined such that $x = \ln(e^x)$. Figure xxx plots e^x and $\ln(x)$. Notice that e^x is always positive and increasing in x. $\ln(x)$ is monotonically increasing in x and is only defined for $x > 0$. Also note that $\ln(1) = 0$ and $\ln(-\infty) = 0$. The exponential and natural logarithm functions have the following properties

1. $\ln(x \cdot y) = \ln(x) + \ln(y)$, $x, y > 0$
2. $\ln(x/y) = \ln(x) - \ln(y)$, $x, y > 0$
3. $\ln(x^y) = y \ln(x)$, $x > 0$
4. $\frac{d \ln(x)}{dx} = \frac{1}{x}$, $x > 0$
5. $\frac{d}{ds} \ln(f(x)) = \frac{1}{f(x)} \frac{d}{dx} f(x)$ (chain-rule)
6. $e^x e^y = e^{x+y}$
7. $e^x e^{-y} = e^{x-y}$
8. $(e^x)^y = e^{xy}$
9. $e^{\ln(x)} = x$
10. $\frac{d}{dx} e^x = e^x$
11. $\frac{d}{dx} e^{f(x)} = e^{f(x)} \frac{d}{dx} f(x)$ (chain-rule)

6 Problems

Exercise 6.1 *Excel exercises*

Go to `http://finance.yahoo.com` and download monthly data on Microsoft (ticker symbol msft) over the period December 1996 to December 2001. See the Project page on the class website for instructions on how to download data from Yahoo. Read the data into Excel and make sure to re-order the data so that time runs forward. Do your analysis on the monthly closing price data (which should be adjusted for dividends and stock splits). Name the spreadsheet tab with the data “data”.

1. Make a time plot (line plot in Excel) of the monthly price data over the period (end of December 1996 through (end of) December 2001. Please put informative titles and labels on the graph. Place this graph in a separate tab (spreadsheet) from the data. Name this tab “graphs”. Comment on what you see (eg. price trends, etc). If you invested \$1,000 at the end of December 1996 what would your investment be worth at the end of December 2001? What is the annual rate of return over this five year period assuming annual compounding?

2. Make a time plot of the natural logarithm of monthly price data over the period December 1986 through December 2000 and place it in the “graph” tab. Comment on what you see and compare with the plot of the raw price data. Why is a plot of the log of prices informative?

3. Using the monthly price data over the period December 1996 through December 2001 in the “data” tab, compute simple (no compounding) monthly returns (Microsoft does not pay a dividend). When computing returns, use the convention that P_t is the end of month closing price. Make a time plot of the monthly returns, place it in the “graphs” tab and comment. Keep in mind that the returns are percent per month and that the annual return on a US T-bill is about 5%.

4. Using the simple monthly returns in the “data” tab, compute simple annual returns for the years 1996 through 2001. Make a time plot of the annual returns, put them in the “graphs” tab and comment. Note: You may compute annual returns using overlapping data or non-overlapping data. With overlapping data you get a series of annual returns for every month (sounds weird, I know). That is, the first month annual return is from the end of December, 1996 to the end of December, 1997. Then second month annual return is from the end of January, 1997 to the end of January, 1998 etc. With non-overlapping data you get a series of 5 annual returns for the 5 year period 1996-2001. That is, the annual return for 1997 is computed from the end of December 1996 through

the end of December 1997. The second annual return is computed from the end of December 1997 through the end of December 1998 etc.

5. Using the monthly price data over the period December 1996 through December 2001, compute continuously compounded monthly returns and place then in the “data” tab. Make a time plot of the monthly returns, put them in the ”graphs” tab and comment. Briefly compare the continuously compounded returns to the simple returns.

6. Using the continuously compounded monthly returns, compute continuously compounded annual returns for the years 1997 through 2001. Make a time plot of the annual returns and comment. Briefly compare the continuously compounded returns to the simple returns.

Exercise 6.2 *Return calculations*

Consider the following (actual) monthly closing price data for Microsoft stock over the period December 1999 through December 2000

End of Month Price Data for Microsoft Stock	
December, 1999	\$116.751
January, 2000	\$97.875
February, 2000	\$89.375
March, 2000	\$106.25
April, 2000	\$69.75
May, 2000	\$62.5625
June, 2000	\$80
July, 2000	\$69.8125
August, 2000	\$69.8125
September, 2000	\$60.3125
October, 2000	\$68.875
November, 2000	\$57.375
December, 2000	\$43.375

1. Using the data in the table, what is the simple monthly return between December, 1999 and January 2000? If you invested \$10,000 in Microsoft at the end of December 1999, how much would the investment be worth at the end of January 2000?

2. Using the data in the table, what is the continuously compounded monthly return between December, 1999 and January 2000? Convert this continuously compounded return to a simple return (you should get the same answer as in part a).

3. Assuming that the simple monthly return you computed in part (1) is the same for 12 months, what is the annual return with monthly compounding?

4. Assuming that the continuously compounded monthly return you computed in part (2) is the same for 12 months, what is the continuously compounded annual return?

5. Using the data in the table, compute the actual simple annual return between December 1999 and December 2000. If you invested $10,000 in Microsoft at the end of December 1999, how much would the investment be worth at the end of December 2000? Compare with your result in part (3).

6. Using the data in the table, compute the actual annual continuously compounded return between December 1999 and December 2000. Compare with your result in part (4). Convert this continuously compounded return to a simple return (you should get the same answer as in part 5).

7 References

References

[1] Campbell, J., A. Lo, and C. MacKinlay (1997), *The Econometrics of Financial Markets*, Princeton University Press.

[2] *Handbook of U.W. Government and Federal Agency Securities and Related Money Market Instruments*, "The Pink Book", 34th ed. (1990), The First Boston Corporation, Boston, MA.

[3] Stigum, M. (1981), *Money Market Calculations: Yields, Break Evens and Arbitrage*, Dow Jones Irwin.

[4] Watsham, T.J. and Parramore, K. (1998), *Quantitative Methods in Finance*, International Thomson Business Press, London, UK.

Introduction to Financial Econometrics Chapter 2 Review of Random Variables and Probability Distributions

Eric Zivot
Department of Economics, University of Washington

January 18, 2000
This version: February 21, 2001

1 Random Variables

We start with a basic definition of a random variable

Definition 1 *A Random variable X is a variable that can take on a given set of values, called the sample space and denoted S_X, where the likelihood of the values in S_X is determined by X's probability distribution function (pdf).*

For example, consider the price of Microsoft stock next month. Since the price of Microsoft stock next month is not known with certainty today, we can consider it a random variable. The price next month must be positive and realistically it can't get too large. Therefore the sample space is the set of positive real numbers bounded above by some large number. It is an open question as to what is the best characterization of the probability distribution of stock prices. The log-normal distribution is one possibility[1].

As another example, consider a one month investment in Microsoft stock. That is, we buy 1 share of Microsoft stock today and plan to sell it next month. Then the return on this investment is a random variable since we do not know its value today with certainty. In contrast to prices, returns can be positive or negative and are bounded from below by -100%. The normal distribution is often a good approximation to the distribution of simple monthly returns and is a better approximation to the distribution of continuously compounded monthly returns.

As a final example, consider a variable X defined to be equal to one if the monthly price change on Microsoft stock is positive and is equal to zero if the price change

[1] If P is a positive random variable such that $\ln P$ is normally distributed the P has a log-normal distribution. We will discuss this distribution is later chapters.

is zero or negative. Here the sample space is trivially the set $\{0, 1\}$. If it is equally likely that the monthly price change is positive or negative (including zero) then the probability that $X = 1$ or $X = 0$ is 0.5.

1.1 Discrete Random Variables

Consider a random variable generically denoted X and its set of possible values or sample space denoted S_X.

De&nition 2 *A discrete random variable X is one that can take on a &nite number of n different values $x_1, x_2, \ldots, x_n$ or, at most, an in&nite number of different values $x_1, x_2, \ldots$.*

De&nition 3 *The pdf of a discrete random variable, denoted $p(x)$, is a function such that $p(x) = \Pr(X = x)$. The pdf must satisfy (i) $p(x) \geq 0$ for all $x \in S_X$; (ii) $p(x) = 0$ for all $x \notin S_X$; and (iii) $\sum_{x \in S_X} p(x) = 1$.*

As an example, let X denote the annual return on Microsoft stock over the next year. We might hypothesize that the annual return will be in! uenced by the general state of the economy. Consider &ve possible states of the economy: depression, recession, normal, mild boom and major boom. A stock analyst might forecast different values of the return for each possible state. Hence X is a discrete random variable that can take on &ve different values. The following table describes such a probability distribution of the return.

Table 1

State of Economy	S_X = Sample Space	$p(x) = \Pr(X = x)$
Depression	-0.30	0.05
Recession	0.0	0.20
Normal	0.10	0.50
Mild Boom	0.20	0.20
Major Boom	0.50	0.05

A graphical representation of the probability distribution is presented in Figure 1.

1.1.1 The Bernoulli Distribution

Let $X = 1$ if the price next month of Microsoft stock goes up and $X = 0$ if the price goes down (assuming it cannot stay the same). Then X is clearly a discrete random variable with sample space $S_X = \{0, 1\}$. If the probability of the stock going up or down is the same then $p(0) = p(1) = 1/2$ and $p(0) + p(1) = 1$.

The probability distribution described above can be given an exact mathematical representation known as the *Bernoulli distribution*. Consider two mutually exclusive events generically called "success" and "failure". For example, a success could be a stock price going up or a coin landing heads and a failure could be a stock price going down or a coin landing tails. In general, let $X = 1$ if success occurs and let $X = 0$ if failure occurs. Let $\Pr(X = 1) = \pi$, where $0 < \pi < 1$, denote the probability of success. Clearly, $\Pr(X = 0) = 1 - \pi$ is the probability of failure. A mathematical model for this set-up is

$$p(x) = \Pr(X = x) = \pi^x(1-\pi)^{1-x},\ x = 0, 1.$$

When $x = 0$, $p(0) = \pi^0(1-\pi)^{1-0} = 1 - \pi$ and when $x = 1, p(1) = \pi^1(1-\pi)^{1-1} = \pi$. This distribution is presented graphically in Figure 2.

1.2 Continuous Random Variables

Definition 4 *A continuous random variable X is one that can take on any real value.*

Definition 5 *The probability density function (pdf) of a continuous random variable X is a nonnegative function p, defined on the real line, such that for any interval A*

$$\Pr(X \in A) = \int_A p(x)dx.$$

That is, $\Pr(X \in A)$ *is the "area under the probability curve over the interval A". The pdf p must satisfy (i) $p(x) \geq 0$; and (ii) $\int_{-\infty}^{\infty} p(x)dx = 1$.*

A typical "bell-shaped" pdf is displayed in Figure 3. In that figure the total area under the curve must be 1, and the value of $\Pr(a \leq X \leq b)$ is equal to the area of the shaded region. For a continuous random variable, $p(x) \neq \Pr(X = x)$ but rather gives the height of the probability curve at x. In fact, $\Pr(X = x) = 0$ for all values of x. That is, probabilities are not defined over single points; they are only defined over intervals.

1.2.1 The Uniform Distribution on an Interval

Let X denote the annual return on Microsoft stock and let a and b be two real numbers such that $a < b$. Suppose that the annual return on Microsoft stock can take on any value between a and b. That is, the sample space is restricted to the interval $S_X = \{x \in \mathcal{R} : a \leq x \leq b\}$. Further suppose that the probability that X will belong to any subinterval of S_X is proportional to the length of the interval. In this case, we say that X is *uniformly distributed* on the interval $[a, b]$. The p.d.f. of X has the very simple mathematical form

$$p(x) = \begin{array}{cc} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{array}$$

and is presented graphically in Figure 4. Notice that the area under the curve over the interval $[a, b]$ integrates to 1 since

$$\int_a^b \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b dx = \frac{1}{b-a} [x]_a^b = \frac{1}{b-a}[b-a] = 1.$$

Suppose, for example, $a = -1$ and $b = 1$ so that $b - a = 2$. Consider computing the probability that the return will be between -50% and 50%.We solve

$$\Pr(-50\% < X < 50\%) = \int_{-0.5}^{0.5} \frac{1}{2} dx = \frac{1}{2} [x]_{-0.5}^{0.5} = \frac{1}{2} [0.5 - (-0.5)] = \frac{1}{2}.$$

Next, consider computing the probability that the return will fall in the interval $[0, \delta]$ where δ is some small number less than $b = 1$:

$$\Pr(0 \leq X \leq \delta) = \frac{1}{2} \int_0^\delta dx = \frac{1}{2} [x]_0^\delta = \frac{1}{2}\delta.$$

As $\delta \to 0$, $\Pr(0 \leq X \leq \delta) \to \Pr(X = 0)$. Using the above result we see that

$$\lim_{\delta \to 0} \Pr(0 \leq X \leq \delta) = \Pr(X = 0) = \lim_{\delta \to 0} \frac{1}{2}\delta = 0.$$

Hence, probabilities are de&ned on intervals but not at distinct points. As a result, for a continuous random variable X we have

$$\Pr(a \leq X \leq b) = \Pr(a \leq X < b) = \Pr(a < X \leq b) = \Pr(a < X < b).$$

1.2.2 The Standard Normal Distribution

The normal or Gaussian distribution is perhaps the most famous and most useful continuous distribution in all of statistics. The shape of the normal distribution is the familiar bell curve. As we shall see, it is also well suited to describe the probabilistic behavior of stock returns.

If a random variable X follows a *standard normal distribution* then we often write $X \sim N(0, 1)$ as short-hand notation. This distribution is centered at zero and has in! ection points at ± 1. The pdf of a normal random variable is given by

$$p(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} \quad -\infty \leq x \leq \infty.$$

It can be shown via the change of variables formula in calculus that the area under the standard normal curve is one:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx = 1.$$

The standard normal distribution is graphed in Figure 5. Notice that the distribution is symmetric about zero; i.e., the distribution has exactly the same form to the left and right of zero.

The normal distribution has the annoying feature that the area under the normal curve cannot be evaluated analytically. That is

$$\Pr(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx$$

does not have a closed form solution. The above integral must be computed by numerical approximation. Areas under the normal curve, in one form or another, are given in tables in almost every introductory statistics book and standard statistical software can be used to &nd these areas. Some useful results from the normal tables are

$$\begin{aligned} \Pr(-1 &< X < 1) \approx 0.67, \\ \Pr(-2 &< X < 2) \approx 0.95, \\ \Pr(-3 &< X < 3) \approx 0.99. \end{aligned}$$

Finding Areas Under the Normal Curve In the back of most introductory statistics textbooks is a table giving information about areas under the standard normal curve. Most spreadsheet and statistical software packages have functions for &nding areas under the normal curve. Let X denote a standard normal random variable. Some tables and functions give $\Pr(0 \leq X < z)$ for various values of $z > 0$, some give $\Pr(X \geq z)$ and some give $\Pr(X \leq z)$. Given that the total area under the normal curve is one and the distribution is symmetric about zero the following results hold:

- $\Pr(X \leq z) = 1 - \Pr(X \geq z)$ and $\Pr(X \geq z) = 1 - \Pr(X \leq z)$
- $\Pr(X \geq z) = \Pr(X \leq -z)$
- $\Pr(X \geq 0) = \Pr(X \leq 0) = 0.5$

The following examples show how to compute various probabilities.

Example 6 *Find* $\Pr(X \geq 2)$. *We know that* $\Pr(X \geq 2) = \Pr(X \geq 0) - \Pr(0 \leq X \leq 2) = 0.5 - \Pr(0 \leq X \leq 2)$. *From the normal tables we have* $\Pr(0 \leq X \leq 2) = 0.4772$ *and so* $\Pr(X \geq 2) = 0.5 - 0.4772 = 0.0228$.

Example 7 *Find* $\Pr(X \leq 2)$. *We know that* $\Pr(X \leq 2) = 1 - \Pr(X \geq 2)$ *and using the result from the previous example we have* $\Pr(X \leq 2) = 1 - 0.0228 = 0.9772$.

Example 8 *Find* $\Pr(-1 \leq X \leq 2)$. *First, note that* $\Pr(-1 \leq X \leq 2) = \Pr(-1 \leq X \leq 0) + \Pr(0 \leq X \leq 2)$. *Using symmetry we have that* $\Pr(-1 \leq X \leq 0) = \Pr(0 \leq X \leq 1) = 0.3413$ *from the normal tables. Using the result from the &rst example we get* $\Pr(-1 \leq X \leq 2) = 0.3413 + 0.4772 = 0.8185$.

1.3 The Cumulative Distribution Function

De&nition 9 *The cumulative distribution function (cdf), F, of a random variable X (discrete or continuous) is simply the probability that $X \leq x$:*

$$F(x) = \Pr(X \leq x), \quad -\infty \leq x \leq \infty.$$

The cdf has the following properties:

- If $x_1 < x_2$ then $F(x_1) \leq F(x_2)$
- $F(-\infty) = 0$ and $F(\infty) = 1$
- $\Pr(X > x) = 1 - F(x)$
- $\Pr(x_1 < X \leq x_2) = F(x_2) - F(x_1)$

The cdf for the discrete distribution of Microsoft is given in Figure 6. Notice that the cdf in this case is a discontinuous step function.

The cdf for the uniform distribution over $[a, b]$ can be determined analytically since

$$F(x) = \Pr(X < x) = \frac{1}{b-a} \int_a^x dt = \frac{1}{b-a} [t]_a^x = \frac{x-a}{b-a}.$$

Notice that for this example, we can determine the pdf of X directly from the cdf via

$$p(x) = F'(x) = \frac{d}{dx} F(x) = \frac{1}{b-a}.$$

The cdf of the standard normal distribution is used so often in statistics that it is given its own special symbol:

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}z^2) dz,$$

where X is a standard normal random variable. The cdf $\Phi(x)$, however, does not have an anaytic representation like the cdf of the uniform distribution and must be approximated using numerical techniques.

1.4 Quantiles of the Distribution of a Random Variable

Consider a random variable X with CDF $F_X(x) = \Pr(X \leq x)$. The $100 \cdot \alpha\%$ *quantile* of the distribution for X is the value q_α that satis&es

$$F_X(q_\alpha) = \Pr(X \leq q_\alpha) = \alpha$$

For example, the 5% quantile of X, $q_{.05}$, satis&es

$$F_X(q_{.05}) = \Pr(X \leq q_{.05}) = .05.$$

The *median* of the distribution is 50% quantile. That is, the *median* satis&es

$$F_X(median) = \Pr(X \leq median) = .5$$

The 5% quantile and the median are illustrated in Figure xxx using the CDF F_X as well as the pdf f_X.

If F_X is invertible then q_a may be determined as

$$q_a = F_X^{-1}(\alpha)$$

where F_X^{-1} denotes the inverse function of F_X. Hence, the 5% quantile and the median may be determined as

$$\begin{aligned} q_{.05} &= F_X^{-1}(.05) \\ median &= F_X^{-1}(.5) \end{aligned}$$

Example 10 *Let $X \tilde{} U[a, b]$ where $b > a$. The cdf of X is given by*

$$\alpha = \Pr(X \leq x) = F_X(x) = \frac{x-a}{b-a}, \ a \leq x \leq b$$

Given α, solving for x gives the inverse cdf

$$x = F_X^{-1}(\alpha) = \ \alpha(b-a) + a, \ 0 \leq \alpha \leq 1$$

Using the inverse cdf, the 5% quantile and median, for example, are given by

$$\begin{aligned} q_{.05} &= F_X^{-1}(.05) = .05(b-a) + a = .05b + .95a \\ median &= F_X^{-1}(.5) = .5(b-a) + a = .5(a+b) \end{aligned}$$

If $a = 0$ and $b = 1$ then $q_{.05} = 0.05$ and $median = 0.5$.

Example 11 *Let $X \tilde{} N(0, 1)$. The quantiles of the standard normal are determined from*

$$q_\alpha = \Phi^{-1}(\alpha)$$

where Φ^{-1} denotes the inverse of the cdf Φ. This inverse function must be approximated numerically. Using the numerical approximation to the inverse function, the 5% quantile and median are given by

$$\begin{aligned} q_{.05} &= \Phi^{-1}(.05) = -1.645 \\ median &= \Phi^{-1}(.5) = 0 \end{aligned}$$

1.5 Shape Characteristics of Probability Distributions

Very often we would like to know certain shape characteristics of a probability distribution. For example, we might want to know where the distribution is centered and how spread out the distribution is about the central value. We might want to know if the distribution is symmetric about the center. For stock returns we might want to know about the likelihood of observing extreme values for returns. This means that we would like to know about the amount of probability in the extreme tails of the distribution. In this section we discuss four shape characteristics of a pdf:

- expected value or mean - center of mass of a distribution
- variance and standard deviation - spread about the mean
- skewness - measure of symmetry about the mean
- kurtosis - measure of “tail thickness”

1.5.1 Expected Value

The expected value of a random variable X, denoted $E[X]$ or μ_X, measures the center of mass of the pdf For a discrete random variable X with sample space S_X

$$\mu_X = E[X] = \sum_{x \in S_X} x \cdot \Pr(X = x).$$

Hence, $E[X]$ is a probability weighted average of the possible values of X.

Example 12 *Using the discrete distribution for the return on Microsoft stock in Table 1, the expected return is*

$$\begin{aligned} E[X] &= (-0.3) \cdot (0.05) + (0.0) \cdot (0.20) + (0.1) \cdot (0.5) + (0.2) \cdot (0.2) + (0.5) \cdot (0.05) \\ &= 0.10. \end{aligned}$$

Example 13 *Let X be a Bernoulli random variable with success probability π. Then*

$$E[X] = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi$$

That is, the expected value of a Bernoulli random variable is its probability of success.

For a continuous random variable X with pdf $p(x)$

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x \cdot p(x) dx.$$

Example 14 *Suppose X has a uniform distribution over the interval $[a, b]$. Then*

$$\begin{aligned} E[X] &= \frac{1}{b-a}\int_a^b x dx = \frac{1}{b-a}\left[\frac{1}{2}x^2\right]_a^b \\ &= \frac{1}{2(b-a)}\left[b^2 - a^2\right] \\ &= \frac{(b-a)(b+a)}{2(b-a)} = \frac{b+a}{2}. \end{aligned}$$

Example 15 *Suppose X has a standard normal distribution. Then it can be shown that*

$$E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0.$$

1.5.2 Expectation of a Function of a Random Variable

The other shape characteristics of distributions are based on expectations of certain functions of a random variable. Let $g(X)$ denote some function of the random variable X. If X is a discrete random variable with sample space S_X then

$$E[g(X)] = \sum_{x \in S_X} g(x) \cdot \Pr(X = x),$$

and if X is a continuous random variable with pdf p then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot p(x) dx.$$

1.5.3 Variance and Standard Deviation

The variance of a random variable X, denoted $var(X)$ or σ_X^2, measures the spread of the distribution about the origin using the function $g(X) = (X - \mu_X)^2$. For a discrete random variable X with sample space S_X

$$\sigma_X^2 = var(X) = E[(X - \mu_X)^2] = \sum_{x \in S_X} (x - \mu_X)^2 \cdot \Pr(X = x).$$

Notice that the variance of a random variable is always nonnegative.

Example 16 *Using the discrete distribution for the return on Microsoft stock in Table 1 and the result that $\mu_X = 0.1$, we have*

$$\begin{aligned} var(X) &= (-0.3 - 0.1)^2 \cdot (0.05) + (0.0 - 0.1)^2 \cdot (0.20) + (0.1 - 0.1)^2 \cdot (0.5) \\ &\quad + (0.2 - 0.1)^2 \cdot (0.2) + (0.5 - 0.1)^2 \cdot (0.05) \\ &= 0.020. \end{aligned}$$

Example 17 *Let X be a Bernoulli random variable with success probability π. Given that $\mu_X = \pi$ it follows that*

$$\begin{aligned} var(X) &= (0-\pi)^2 \cdot (1-\pi) + (1-\pi)^2 \cdot \pi \\ &= \pi^2(1-\pi) + (1-\pi^2)\pi \\ &= \pi(1-\pi)\left[\pi + (1-\pi)\right] \\ &= \pi(1-\pi). \end{aligned}$$

The standard deviation of X, denoted $SD(X)$ or σ_X, is just the square root of the variance. Notice that $SD(X)$ is in the same units of measurement as X whereas $var(X)$ is in squared units of measurement. For bell-shaped or normal looking distributions the SD measures the typical size of a deviation from the mean value.

Example 18 *For the distribution in Table 1, we have $SD(X) = \sigma_X = \sqrt{0.020} = 0.141$. Given that the distribution is fairly bell-shaped we can say that typical values deviate from the mean value of 10% by about 14.1%.*

For a continuous random variable X with pdf $p(x)$

$$\sigma_X^2 = var(X) = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty} (x-\mu_X)^2 \cdot p(x)dx.$$

Example 19 *Suppose X has a standard normal distribution so that $\mu_X = 0$. Then it can be shown that*

$$var\,(X) = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1,$$

and so $SD(X) = 1$.

1.5.4 The General Normal Distribution

Recall, if X has a standard normal distribution then $E[X] = 0$, $var(X) = 1$. If X has general normal distribution, denoted $X \sim N(\mu_X, \sigma_X^2)$, then its pdf is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2}, \quad -\infty \leq x \leq \infty.$$

It can be shown that $E[X] = \mu_X$ and $var(X) = \sigma_X^2$, although showing these results analytically is a bit of work and is good calculus practice. As with the standard normal distribution, areas under the general normal curve cannot be computed analytically. Using numerical approximations, it can be shown that

$$\begin{aligned} \Pr(\mu_X - \sigma_X &< X < \mu_X + \sigma_X) \approx 0.67, \\ \Pr(\mu_X - 2\sigma_X &< X < \mu_X + 2\sigma_X) \approx 0.95, \\ \Pr(\mu_X - 3\sigma_X &< X < \mu_X + 3\sigma_X) \approx 0.99. \end{aligned}$$

Hence, for a general normal random variable about 95% of the time we expect to see values within ± 2 standard deviations from its mean. Observations more than three standard deviations from the mean are very unlikely.

(insert &gures showing different normal distributions)

1.5.5 The Log-Normal distribution

A random variable Y is said to be log-normally distributed with parameters μ and σ^2 if

$$\ln Y \,\tilde{}\, N(\mu, \sigma^2).$$

Equivalently, let $X \,\tilde{}\, N(\mu, \sigma^2)$ and de&ne

$$Y = e^X.$$

Then Y is log-normally distributed and is denoted $Y \,\tilde{}\, \ln N(\mu, \sigma^2)$.

(insert &gure showing lognormal distribution).

It can be shown that

$$\begin{aligned} \mu_Y &= E[Y] = e^{\mu+\sigma^2/2} \\ \sigma_Y^2 &= var(Y) = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) \end{aligned}$$

Example 20 *Let $r_t = \ln(P_t/P_{t-1})$ denote the continuously compounded monthly return on an asset and assume that $r_t \,\tilde{}\, N(\mu, \sigma^2)$. Let $R_t = \frac{P_t - P_{t-1}}{P_t}$ denote the simple monthly return. The relationship between r_t and R_t is given by $r_t = \ln(1 + R_t)$ and $1+R_t = e^{r_t}$. Since r_t is normally distributed $1+R_t$ is log-normally distributed. Notice that the distribution of $1 + R_t$ is only de&ned for positive values of $1 + R_t$. This is appropriate since the smallest value that R_t can take on is -1.*

1.5.6 Using standard deviation as a measure of risk

Consider the following investment problem. We can invest in two non-dividend paying stocks A and B over the next month. Let R_A denote monthly return on stock A and R_B denote the monthly return on stock B. These returns are to be treated as random variables since the returns will not be realized until the end of the month. We assume that $R_A \,\tilde{}\, N(\mu_A, \sigma_A^2)$ and $R_B \,\tilde{}\, N(\mu_B, \sigma_B^2)$. Hence, μ_i gives the expected return, $E[R_i]$, on asset i and σ_i gives the typical size of the deviation of the return on asset i from its expected value. Figure xxx shows the pdfs for the two returns. Notice that $\mu_A > \mu_B$ but also that $\sigma_A > \sigma_B$. The return we expect on asset A is bigger than the return we expect on asset B but the variability of the return on asset A is also greater than the variability on asset B. The high return variability of asset A re! ects the risk associated with investing in asset A. In contrast, if we invest in asset B we get a lower expected return but we also get less return variability or risk. This example illustrates the fundamental "no free lunch" principle of economics and finance: you can't get something for nothing. In general, to get a higher return you must take on extra risk.

1.5.7 Skewness

The skewness of a random variable X, denoted $skew(X)$, measures the symmetry of a distribution about its mean value using the function $g(X) = (X - \mu_X)^3/\sigma_X^3$, where σ_X^3 is just $SD(X)$ raised to the third power. For a discrete random variable X with sample space S_X

$$skew(X) = \frac{E[(X-\mu_X)^3]}{\sigma_X^3} = \frac{\sum_{x \in S_X}(x-\mu_X)^3 \cdot \Pr(X=x)}{\sigma_X^3}.$$

If X has a symmetric distribution then $skew(X) = 0$ since positive and negative values in the formula for skewness cancel out. If $skew(X) > 0$ then the distribution of X has a "long right tail" and if $skew(X) < 0$ the distribution of X has a "long left tail". These cases are illustrated in Figure 6.

Example 21 *Using the discrete distribution for the return on Microsoft stock in Table 1, the results that $\mu_X = 0.1$ and $\sigma_X = 0.141$, we have*

$$\begin{aligned} skew(X) &= [(-0.3-0.1)^3 \cdot (0.05) + (0.0-0.1)^3 \cdot (0.20) + (0.1-0.1)^3 \cdot (0.5) \\ &\quad +(0.2-0.1)^3 \cdot (0.2) + (0.5-0.1)^3 \cdot (0.05)]/(0.141)^3 \\ &= 0.0 \end{aligned}$$

For a continuous random variable X with pdf $p(x)$

$$skew(X) = \frac{E[(X-\mu_X)^3]}{\sigma_X^3} = \frac{\int_{-\infty}^{\infty}(x-\mu_X)^3 \cdot p(x)dx}{\sigma_X^3}.$$

Example 22 *Suppose X has a general normal distribution with mean μ_X and variance σ_X^2. Then it can be shown that*

$$skew(X) = \int_{-\infty}^{\infty} \frac{(x-\mu_X)^3}{\sigma_X^3} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2} dx = 0.$$

This result is expected since the normal distribution is symmetric about it's mean value μ_X.

1.5.8 Kurtosis

The kurtosis of a random variable X, denoted $kurt(X)$, measures the thickness in the tails of a distribution and is based on $g(X) = (X - \mu_X)^4/\sigma_X^4$. For a discrete random variable X with sample space S_X

$$kurt(X) = \frac{E[(X-\mu_X)^4]}{\sigma_X^4} = \frac{\sum_{x \in S_X}(x-\mu_X)^2 \cdot \Pr(X=x)}{\sigma_X^4},$$

where σ_X^4 is just $SD(X)$ raised to the fourth power. Since kurtosis is based on deviations from the mean raised to the fourth power, large deviations get lots of weight. Hence, distributions with large kurtosis values are ones where there is the possibility of extreme values. In contrast, if the kurtosis is small then most of the observations are tightly clustered around the mean and there is very little probability of observing extreme values.

Example 23 *Using the discrete distribution for the return on Microsoft stock in Table 1, the results that $\mu_X = 0.1$ and $\sigma_X = 0.141$, we have*

$$\begin{aligned} kurt(X) &= [(-0.3-0.1)^4 \cdot (0.05) + (0.0-0.1)^4 \cdot (0.20) + (0.1-0.1)^4 \cdot (0.5) \\ &\quad +(0.2-0.1)^4 \cdot (0.2) + (0.5-0.1)^4 \cdot (0.05)]/(0.141)^4 \\ &= 6.5 \end{aligned}$$

For a continuous random variable X with pdf $p(x)$

$$kurt(X) = \frac{E[(X-\mu_X)^4]}{\sigma_X^4} = \frac{\int_{-\infty}^{\infty}(x-\mu_X)^4 \cdot p(x)dx}{\sigma_X^4}.$$

Example 24 *Suppose X has a general normal distribution mean μ_X and variance σ_X^2. Then it can be shown that*

$$kurt(X) = \int_{-\infty}^{\infty} \frac{(x-\mu_X)^4}{\sigma_X^4} \cdot \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{1}{2}(x-\mu_X)^2} dx = 3.$$

Hence a kurtosis of 3 is a benchmark value for tail thickness of bell-shaped distributions. If a distribution has a kurtosis greater than 3 then the distribution has thicker tails than the normal distribution and if a distribution has kurtosis less than 3 then the distribution has thinner tails than the normal.

Sometimes the kurtosis of a random variable is described relative to the kurtosis of a normal random variable. This relative value of kurtosis is referred to as *excess kurtosis* and is de&ned as

$$excess\ kurt(X) = kurt(X) - 3$$

If excess the excess kurtosis of a random variable is equal to zero then the random variable has the same kurtosis as a normal random variable. If excess kurtosis is greater than zero, then kurtosis is larger than that for a normal; if excess kurtosis is less than zero, then kurtosis is less than that for a normal.

1.6 Linear Functions of a Random Variable

Let X be a random variable either discrete or continuous with $E[X] = \mu_X$, $var(X) = \sigma_X^2$ and let a and b be known constants. De&ne a new random variable Y via the linear function of X

$$Y = g(X) = aX + b.$$

Then the following results hold:

- $E[Y] = aE[X] + b$ or $\mu_Y = a\mu_X + b$.
- $var(Y) = a^2 var(X)$ or $\sigma_Y^2 = a^2\sigma_X^2$.

The &rst result shows that expectation is a linear operation. That is,

$$E[aX + b] = aE[X] + b.$$

In the second result notice that adding a constant to X does not affect its variance and that the effect of multiplying X by the constant a increases the variance of X by the square of a. These results will be used often enough that it useful to go through the derivations, at least for the case that X is a discrete random variable.

Proof. Consider the &rst result. By the de&nition of $E[g(X)]$ with $g(X) = b + aX$ we have

$$\begin{aligned}
E[Y] &= \sum_{x \in S_X} (ax + b) \cdot \Pr(X = x) \\
&= a \sum_{x \in S_X} x \cdot \Pr(X = x) + b \sum_{x \in S_X} \Pr(X = x) \\
&= aE[X] + b \cdot 1 \\
&= a\mu_X + b \\
&= \mu_Y.
\end{aligned}$$

Next consider the second result. Since $\mu_Y = a\mu_X + b$ we have

$$\begin{aligned}
var(Y) &= E[(Y - \mu_y)^2] \\
&= E[(aX + b - (a\mu_X + b))^2] \\
&= E[(a(X - \mu_X) + (b - b))^2] \\
&= E[a^2(X - \mu_X)^2] \\
&= a^2 E[(X - \mu_X)^2] \quad \text{(by the linearity of } E[\cdot]) \\
&= a^2 var(X) \\
& \quad a^2\sigma_X^2.
\end{aligned}$$

Notice that our proof of the second result works for discrete and continuous random variables. ■

A normal random variable has the special property that a linear function of it is also a normal random variable. The following proposition establishes the result.

Proposition 25 *Let* $X \sim N(\mu_X, \sigma_X^2)$ *and let* a *and* b *be constants. Let* $Y = aX + b$. *Then* $Y \sim N(a\mu_X + b, a^2\sigma_X^2)$.

The above property is special to the normal distribution and may or may not hold for a random variable with a distribution that is not normal.

1.6.1 Standardizing a Random Variable

Let X be a random variable with $E[X] = \mu_X$ and $var(X) = \sigma_X^2$. Define a new random variable Z as

$$Z = \frac{X - \mu_X}{\sigma_X} = \frac{1}{\sigma_X}X - \frac{\mu_X}{\sigma_X}$$

which is a linear function $aX + b$ where $a = \frac{1}{\sigma_X}$ and $b = -\frac{\mu_X}{\sigma_X}$. This transformation is called "standardizing" the random variable X since, using the results of the previous section,

$$\begin{aligned} E[Z] &= \frac{1}{\sigma_X}E[X] - \frac{\mu_X}{\sigma_X} = \frac{1}{\sigma_X}\mu_X - \frac{\mu_X}{\sigma_X} = 0 \\ var(Z) &= \left(\frac{1}{\sigma_X}\right)^2 var(X) = \frac{\sigma_X^2}{\sigma_X^2} = 1. \end{aligned}$$

Hence, standardization creates a new random variable with mean zero and variance 1. In addition, if X is normally distributed then so is Z.

Example 26 *Let* $X \sim N(2, 4)$ *and suppose we want to find* $\Pr(X > 5)$. *Since* X *is not standard normal we can't use the standard normal tables to evaluate* $\Pr(X > 5)$ *directly. We solve the problem by standardizing* X *as follows:*

$$\begin{aligned} \Pr(X > 5) &= \Pr\left(\frac{X - 2}{\sqrt{4}} > \frac{5 - 2}{\sqrt{4}}\right) \\ &= \Pr\left(Z > \frac{3}{2}\right) \end{aligned}$$

where $Z \sim N(0, 1)$ *is the standardized value of* X. $\Pr\left(Z > \frac{3}{2}\right)$ *can be found directly from the standard normal tables.*

Standardizing a random variable is often done in the construction of test statistics. For example, the so-called "t-statistic" or "t-ratio" used for testing simple hypotheses on coefficients in the linear regression model is constructed by the above standardization process.

A non-standard random variable X with mean μ_X and variance σ_X^2 can be created from a standard random variable via the linear transformation

$$X = \mu_X + \sigma_X Z.$$

This result is useful for modeling purposes. For example, in Chapter 3 we will consider the Constant Expected Return (CER) model of asset returns. Let R denote the monthly continuously compounded return on an asset and let $\mu = E[R]$ and $\sigma^2 = var(R)$. A simpli&ed version of the CER model is

$$R = \mu + \sigma \cdot \varepsilon$$

where ε is a random variable with mean zero and variance 1. The random variable ε is often interpreted as representing the random news arriving in a given month that makes the observed return differ from the expected value μ. The fact that ε has mean zero means that new, on average, is neutral. The value of σ represents the typical size of a news shock.

(Stuff to add: General functions of a random variable and the change of variables formula. Example with the log-normal distribution)

1.7 Value at Risk

To illustrate the concept of Value-at-Risk (VaR), consider an investment of \$10,000 in Microsoft stock over the next month. Let R denote the monthly *simple* return on Microsoft stock and assume that $R\ \tilde{}N(0.05, (0.10)^2)$. That is, $E[R] = \mu = 0.05$ and $var(R) = \sigma^2 = (0.10)^2$. Let W_0 denote the investment value at the beginning of the month and W_1 denote the investment value at the end of the month. In this example, $W_0 = \$10,000$. Consider the following questions:

- What is the probability distribution of end of month wealth, W_1?
- What is the probability that end of month wealth is less than $\$9,000$ and what must the return on Microsoft be for this to happen?
- What is the monthly VaR on the $\$10,000$ investment in Microsoft stock with 5% probability? That is, what is the loss that would occur if the return on Microsoft stock is equal to its 5% quantile, $q_{.05}$?

To answer the &rst question, note that end of month wealth W_1 is related to initial wealth W_0 and the return on Microsoft stock R via the linear function

$$\begin{aligned} W_1 &= W_0(1+R) = W_0 + W_0R \\ &= \$10,000 + \$10,000 \cdot R. \end{aligned}$$

Using the properties of linear functions of a random variable we have

$$\begin{aligned} E[W_1] &= W_0 + W_0E[R] \\ &= \$10,000 + \$10,000(0.05) = \$10,500 \end{aligned}$$

and

$$\begin{aligned} var(W_1) &= (W_0)^2 var(R) \\ &= (\$10,000)^2(0.10)^2, \\ SD(W_1) &= (\$10,000)(0.10) = \$1,000. \end{aligned}$$

Further, since R is assumed to be normally distributed we have

$$W_1 \sim N(\$10,500, (\$1,000)^2)$$

To answer the second question, we use the above normal distribution for W_1 to get

$$\Pr(W_1 < \$9,000) = 0.067$$

To &nd the return that produces end of month wealth of \$9,000 or a loss of $\$10,000 - \$9,000 = \$1,000$ we solve

$$R^* = \frac{\$9,000 - \$10,000}{\$10,000} = -0.10.$$

In other words, if the monthly return on Microsoft is -10% or less then end of month wealth will be \$9,000 or less. Notice that -0.10 is the 6.7% quantile of the distribution of R :

$$\Pr(R < -0.10) = 0.067$$

The third question can be answered in two equivalent ways. First, use $R \sim N(0.05, (0.10)^2)$ and solve for the the 5% quantile of Microsoft Stock:

$$\Pr(R < q_{.05}^R) = 0.05 \Rightarrow q_{.05}^R = -0.114.$$

That is, with 5% probability the return on Microsoft stock is -11.4% or less. Now, if the return on Microsoft stock is -11.4% the loss in investment value is $\$10,000 \cdot (0.114) = \$1,144$. Hence, \$1,144 is the 5% VaR over the next month on the \$10,000 investment in Microsoft stock. In general, if W_0 represents the initial wealth and $q_{.05}^R$ is the 5% quantile of distribution of R then the 5% VaR is

$$5\% \text{ VaR } = |W_0 \cdot q_{.05}^R|.$$

For the second method, use $W_1 \sim N(\$10,500, (\$1,000)^2)$ and solve for the 5% quantile of end of month wealth:

$$\Pr(W_1 < q_{.05}^{W_1}) = 0.05 \Rightarrow q_{.05}^{W_1} = \$8,856$$

This corresponds to a loss of investment value of $\$10,000 - \$8,856 = \$1,144$. Hence, if W_0 represents the initial wealth and $q_{.05}^{W_1}$ is the 5% quantile of the distribution of W_1 then the 5% VaR is

$$5\% \text{ VaR} = W_0 - q_{.05}^{W_1}.$$

(insert VaR calculations based on continuously compounded returns)

1.8 Log-Normal Distribution and Jensen's Inequality

(discuss Jensen's inequality: $E[g(X)] < g(E[X])$ for a convex function. Use this to illustrate the difference between $E[W_0 \exp(R)]$ and $W_0 \exp(E[R])$ where R is a continuously compounded return.) Note, this is where the log-normal distribution will come in handy.

2 Bivariate Distributions

So far we have only considered probability distributions for a single random variable. In many situations we want to be able to characterize the probabilistic behavior of two or more random variables simultaneously.

2.1 Discrete Random Variables

For example, let X denote the monthly return on Microsoft Stock and let Y denote the monthly return on Apple computer. For simplicity suppose that the sample spaces for X and Y are $S_X = \{0, 1, 2, 3\}$ and $S_Y = \{0, 1\}$ so that the random variables X and Y are discrete. The joint sample space is the two dimensional grid $S_{XY} = \{(0, 0), (0, 1), (1, 0), (1, 1), (2, 0), (2, 1), (3, 0), (3, 1)\}$. The likelihood that X and Y takes values in the joint sample space is determined by the joint probability distribution

$$p(x, y) = \Pr(X = x, Y = y).$$

The function $p(x, y)$ satisfies

(i) $p(x, y) > 0$ for $x, y \in S_{XY}$;

(ii) $p(x, y) = 0$ for $x, y \notin S_{XY}$;

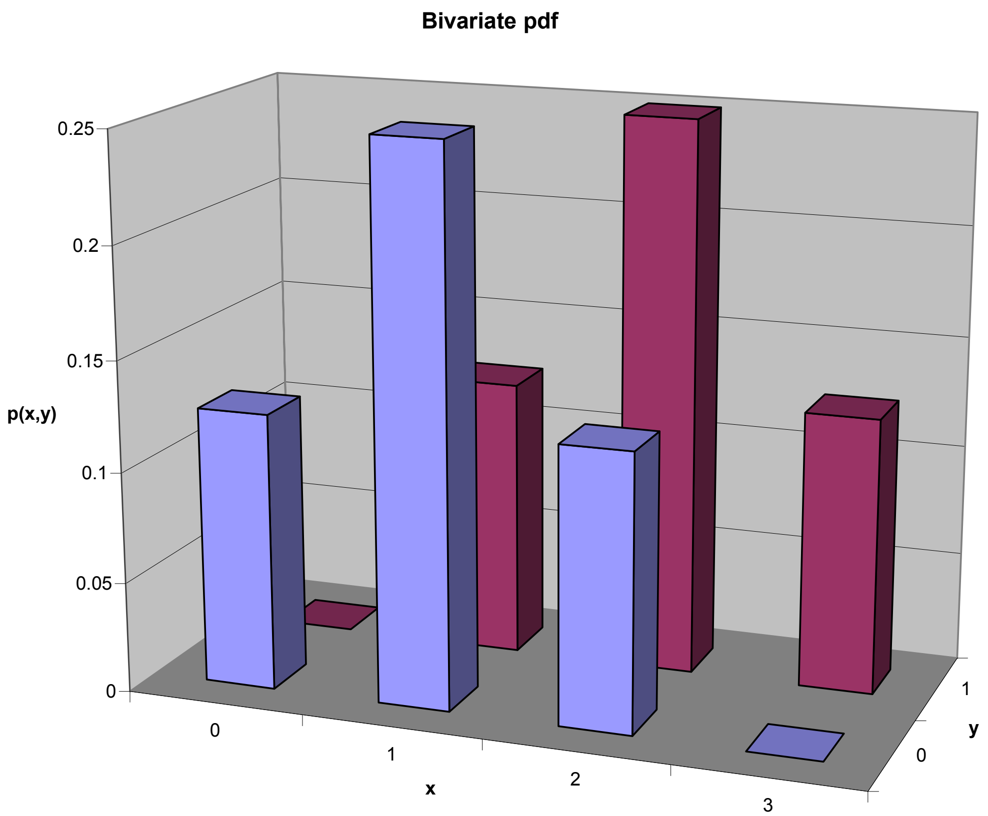
(iii) $\sum_{x,y \in S_{XY}} p(x, y) = \sum_{x \in S_X} \sum_{y \in S_Y} p(x, y) = 1$.

Table 2 illustrates the joint distribution for X and Y.

Table 2

		Y		
	%	0	1	$\Pr(X)$
	0	1/8	0	1/8
X	1	2/8	1/8	3/8
	2	1/8	2/8	3/8
	3	0	1/8	1/8
	$\Pr(Y)$	4/8	4/8	1

For example, $p(0,0) = \Pr(X = 0, Y = 0) = 1/8$. Notice that sum of all the entries in the table sum to unity. The bivariate distribution is illustrated graphically in Figure xxx.



2.1.1 Marginal Distributions

What if we want to know only about the likelihood of X occurring? For example, what is $\Pr(X = 0)$ regardless of the value of Y? Now X can occur if $Y = 0$ or if $Y = 1$ and since these two events are mutually exclusive we have that $\Pr(X = 0) = \Pr(X = 0, Y = 0) + \Pr(X = 0, Y = 1) = 0 + 1/8 = 1/8$. Notice that this probability is equal to the horizontal (row) sum of the probabilities in the table at $X = 0$. The probability $\Pr(X = x)$ is called the *marginal probability* of X and is given by

$$\Pr(X = x) = \sum_{y \in S_Y} \Pr(X = x, Y = y).$$

The marginal probabilities of $X = x$ are given in the last column of Table 2. Notice that the marginal probabilities sum to unity.

We can find the marginal probability of Y in a similar fashion. For example, using the data in Table 2 $\Pr(Y = 1) = \Pr(X = 0, Y = 1) + \Pr(X = 1, Y = 1) + \Pr(X = 2, Y = 1) + \Pr(X = 3, Y = 1) = 0 + 1/8 + 2/8 + 1/8 = 4/8$. This probability is the vertical (column) sum of the probabilities in the table at $Y = 1$. Hence, the marginal probability of $Y = y$ is given by

$$\Pr(Y = y) = \sum_{x \in S_X} \Pr(X = x, Y = y).$$

The marginal probabilities of $Y = y$ are given in the last row of Table 2. Notice that these probabilities sum to 1.

For future reference we note that

$$\begin{aligned} E[X] &= xx, var(X) = xx \\ E[Y] &= xx, var(Y) = xx \end{aligned}$$

2.2 Conditional Distributions

Suppose we know that the random variable Y takes on the value $Y = 0$. How does this knowledge affect the likelihood that X takes on the values $0, 1, 2$ or 3? For example, what is the probability that $X = 0$ *given that* we know $Y = 0$? To find this probability, we use Bayes' law and compute the *conditional probability*

$$\Pr(X = 0|Y = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(Y = 0)} = \frac{1/8}{4/8} = 1/4.$$

The notation $\Pr(X = 0|Y = 0)$ is read as "the probability that $X = 0$ given that $Y = 0$". Notice that the conditional probability that $X = 0$ given that $Y = 0$ is greater than the marginal probability that $X = 0$. That is, $\Pr(X = 0|Y = 0) = 1/4 > \Pr(X = 0) = 1/8$. Hence, knowledge that $Y = 0$ increases the likelihood that $X = 0$. Clearly, X *depends* on Y.

Now suppose that we know that $X = 0$. How does this knowledge affect the probability that $Y = 0$? To find out we compute

$$\Pr(Y = 0|X = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(X = 0)} = \frac{1/8}{1/8} = 1.$$

Notice that $\Pr(Y = 0|X = 0) = 1 > \Pr(Y = 0) = 1/2$. That is, knowledge that $X = 0$ makes it certain that $Y = 0$.

In general, the conditional probability that $X = x$ given that $Y = y$ is given by

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

and the conditional probability that $Y = y$ given that $X = x$ is given by

$$\Pr(Y = y|X = x) = \frac{\Pr(X = x, Y = y)}{\Pr(X = x)}.$$

For the example in Table 2, the conditional probabilities along with marginal probabilities are summarized in Tables 3 and 4. The conditional and marginal distributions of X are graphically displayed in &gure xxx and the conditional and marginal distribution of Y are displayed in &gure xxx. Notice that the marginal distribution of X is centered at $x = 3/2$ whereas the conditional distribution of $X|Y = 0$ is centered at $x = 1$ and the conditional distribution of $X|Y = 1$ is centered at $x = 2$.

Table 3

x	$\Pr(X = x)$	$\Pr(X\|Y = 0)$	$\Pr(X\|Y = 1)$
0	1/8	2/8	0
1	3/8	4/8	2/8
2	3/8	2/8	4/8
3	1/8	0	2/8

Table 4

y	$\Pr(Y = y)$	$\Pr(Y\|X = 0)$	$\Pr(Y\|X = 1)$	$\Pr(Y\|X = 2)$	$\Pr(Y\|X = 3)$
0	1/2	1	2/3	1/3	0
1	1/2	0	1/3	2/3	1

2.2.1 Conditional Expectation and Conditional Variance

Just as we de&ned shape characteristics of the marginal distributions of X and Y we can also de&ne shape characteristics of the conditional distributions of $X|Y = y$ and $Y|X = x$. The most important shape characteristics are the *conditional expectation* (*conditional mean*) and the *conditional variance*. The conditional mean of $X|Y = y$ is denoted by $\mu_{X|Y=y} = E[X|Y = y]$ and the conditional mean of $Y|X = x$ is denoted by $\mu_{Y|X=x} = E[Y|X = x]$. These means are computed as

$$\begin{aligned} \mu_{X|Y=y} &= E[X|Y = y] = \sum_{x \in S_X} x \cdot \Pr(X = x|Y = y), \\ \mu_{Y|X=x} &= E[Y|X = x] = \sum_{y \in S_Y} y \cdot \Pr(Y = y|X = x). \end{aligned}$$

Similarly, the conditional variance of $X|Y = y$ is denoted by $\sigma^2_{X|Y=y} = var(X|Y = y)$ and the conditional variance of $Y|X = x$ is denoted by $\sigma^2_{Y|X=x} = var(Y|X = x)$. These variances are computed as

$$\begin{aligned} \sigma^2_{X|Y=y} &= var(X|Y = y) = \sum_{x \in S_X} (x - \mu_{X|Y=y})^2 \cdot \Pr(X = x|Y = y), \\ \sigma^2_{Y|X=x} &= var(Y|X = x) = \sum_{y \in S_Y} (y - \mu_{Y|X=x})^2 \cdot \Pr(Y = y|X = x). \end{aligned}$$

Example 27 *For the data in Table 2, we have*

$$\begin{aligned}
E[X|Y &= 0] = 0\cdot 1/4 + 1\cdot 1/2 + 2\cdot 1/4 + 3\cdot 0 = 1 \\
E[X|Y &= 1] = 0\cdot 0 + 1\cdot 1/4 + 2\cdot 1/2 + 3\cdot 1/4 = 2 \\
var(X|Y &= 0) = (0-1)^2\cdot 1/4 + (1-1)^2\cdot 1/2 + (2-1)^2\cdot 1/2 + (3-1)^2\cdot 0 = 1/2 \\
var(X|Y &= 1) = (0-2)^2\cdot 0 + (1-2)^2\cdot 1/4 + (2-2)^2\cdot 1/2 + (3-2)^2\cdot 1/4 = 1/2.
\end{aligned}$$

Using similar calculations gives

$$\begin{aligned}
E[Y|X &= 0] = 0, E[Y|X=1] = 1/3, E[Y|X=2] = 2/3, E[Y|X=3] = 1 \\
var(Y|X &= 0) = 0, var(Y|X=1) = 0, var(Y|X=2) = 0, var(Y|X=3) = 0.
\end{aligned}$$

2.2.2 Conditional Expectation and the Regression Function

Consider the problem of predicting the value Y given that we know $X = x$. A natural predictor to use is the conditional expectation $E[Y|X = x]$. In this prediction context, the conditional expectation $E[Y|X = x]$ is called the *regression function*. The graph with $E[Y|X = x]$ on the verticle axis and x on the horizontal axis gives the so-called *regression line*. The relationship between Y and the regression function may expressed using the trivial identity

$$\begin{aligned}
Y &= E[Y|X=x] + Y - E[Y|X=x] \\
&= E[Y|X=x] + \varepsilon
\end{aligned}$$

where $\varepsilon = Y - E[Y|X]$ is called the *regression error*.

Example 28 *For the data in Table 2, the regression line is plotted in figure xxx. Notice that there is a linear relationship between $E[Y|X = x]$ and x. When such a linear relationship exists we call the regression function a linear regression. It is important to stress that linearity of the regression function is not guaranteed.*

2.2.3 Law of Total Expectations

Notice that

$$\begin{aligned}
E[X] &= E[X|Y=0]\cdot \Pr(Y=0) + E[X|Y=1]\cdot \Pr(Y=1) \\
&= 1\cdot 1/2 + 2\cdot 1/2 = 3/2
\end{aligned}$$

and

$$\begin{aligned}
E[Y] &= E[Y|X=0]\cdot \Pr(X=0) + E[Y|X=1]\cdot \Pr(X=1) + E[Y|X=2]\cdot \Pr(X=2) + E[Y|X=3 \\
&= 1/2
\end{aligned}$$

This result is known as the *law of total expectations*. In general, for two random variables X and Y we have

$$\begin{aligned}
E[X] &= E[E[X|Y]] \\
E[Y] &= E[E[Y|X]]
\end{aligned}$$

2.3 Bivariate Distributions for Continuous Random Variables

Let X and Y be continuous random variables defined over the real line. We characterize the joint probability distribution of X and Y using the joint probability function (pdf) $p(x, y)$ such that $p(x, y) \geq 0$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1.$$

For example, in Figure xxx we illustrate the pdf of X and Y as a "bell-shaped" surface in two dimensions. To compute joint probabilities of $x_1 \leq X \leq x_2$ and $y_1 \leq Y \leq y_2$ we need to find the volume under the probability surface over the grid where the intervals $[x_1, x_2]$ and $[y_1, y_2]$ overlap. To find this volume we must solve the double integral

$$\Pr(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) dx dy.$$

Example 29 *A standard bivariate normal pdf for X and Y has the form*

$$p(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}, -\infty \leq x, y \leq \infty$$

and has the shape of a symmetric bell centered at $x = 0$ and $y = 0$ as illustrated in Figure xxx (insert figure here). To find $\Pr(-1 < X < 1, -1 < Y < 1)$ *we must solve*

$$\int_{-1}^{1} \int_{-1}^{1} \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} dx dy$$

which, unfortunately, does not have an analytical solution. Numerical approximation methods are required to evaluate the above integral.

2.3.1 Marginal and Conditional Distributions

The marginal pdf of X is found by integrating y out of the joint pdf $p(x, y)$ and the marginal pdf of Y is found by integrating x out of the joint pdf:

$$\begin{aligned} p(x) &= \int_{-\infty}^{\infty} p(x, y) dy, \\ p(y) &= \int_{-\infty}^{\infty} p(x, y) dx. \end{aligned}$$

The conditional pdf of X given that $Y = y$, denoted $p(x|y)$, is computed as

$$p(x|y) = \frac{p(x, y)}{p(y)}$$

and the conditional pdf of Y given that $X = x$ is computed as

$$p(y|x) = \frac{p(x,y)}{p(x)}.$$

The conditional means are computed as

$$\begin{aligned}
\mu_{X|Y=y} &= E[X|Y=y] = \int x \cdot p(x|y)dx, \\
\mu_{Y|X=x} &= E[Y|X=x] = \int y \cdot p(y|x)dy
\end{aligned}$$

and the conditional variances are computed as

$$\begin{aligned}
\sigma^2_{X|Y=y} &= var(X|Y=y) = \int (x - \mu_{X|Y=y})^2 p(x|y)dx, \\
\sigma^2_{Y|X=x} &= var(Y|X=x) = \int (y - \mu_{Y|X=x})^2 p(y|x)dy.
\end{aligned}$$

2.4 Independence

Let X and Y be two random variables. Intuitively, X is independent of Y if knowledge about Y does not in! uence the likelihood that $X = x$ for all possible values of $x \in S_X$ and $y \in S_Y$. Similarly, Y is independent of X if knowledge about X does not in! uence the likelihood that $Y = y$ for all values of $y \in S_Y$. We represent this intuition formally for discrete random variables as follows.

De&nition 30 *Let X and Y be discrete random variables with sample spaces S_X and S_Y, respectively. X and Y are independent random variables iff*

$$\begin{aligned}
\Pr(X &= x|Y=y) = \Pr(X=x), \text{ for all } x \in S_X, y \in S_Y \\
\Pr(Y &= y|X=x) = \Pr(Y=y), \text{ for all } x \in S_X, y \in S_Y
\end{aligned}$$

Example 31 *For the data in Table 2, we know that* $\Pr(X = 0|Y = 0) = 1/4 \neq \Pr(X = 0) = 1/8$ *so X and Y are not independent.*

Proposition 32 *Let X and Y be discrete random variables with sample spaces S_X and S_Y, respectively. If X and Y are independent then*

$$\Pr(X = x, Y = y) = \Pr(X = x) \cdot \Pr(Y = y), \text{ for all } x \in S_X, y \in S_Y$$

For continuous random variables, we have the following de&nition of independence

De&nition 33 *Let X and Y be continuous random variables. X and Y are independent iff*

$$\begin{aligned}
p(x|y) &= p(x), \text{ for } -\infty < x, y < \infty \\
p(y|x) &= p(y), \text{ for } -\infty < x, y < \infty
\end{aligned}$$

Proposition 34 *Let X and Y be continuous random variables . X and Y are independent iff*

$$p(x, y) = p(x)p(y)$$

The result in the proposition is extremely useful because it gives us an easy way to compute the joint pdf for two independent random variables: we simple compute the product of the marginal distributions.

Example 35 *Let $X \sim N(0, 1)$, $Y \sim N(0, 1)$ and let X and Y be independent. Then*

$$p(x, y) = p(x)p(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2} = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}.$$

This result is a special case of the bivariate normal distribution.

(stuff to add: if X and Y are independent then $f(X)$ and $g(Y)$ are independent for any functions $f(\cdot)$ and $g(\cdot)$.)

2.5 Covariance and Correlation

Let X and Y be two discrete random variables. Figure xxx displays several bivariate probability scatterplots (where equal probabilities are given on the dots).

(insert &gure here)

In panel (a) we see no linear relationship between X and Y. In panel (b) we see a perfect positive linear relationship between X and Y and in panel (c) we see a perfect negative linear relationship. In panel (d) we see a positive, but not perfect, linear relationship. Finally, in panel (e) we see no systematic linear relationship but we see a strong nonlinear (parabolic) relationship. The *covariance* between X and Y measures the *direction* of linear relationship between the two random variables. The *correlation* between X and Y measures the *direction* and *strength* of linear relationship between the two random variables.

Let X and Y be two random variables with $E[X] = \mu_X$, $var(X) = \sigma_X^2$, $E[Y] = \mu_Y$ and $var(Y) = \sigma_Y^2$.

De&nition 36 *The covariance between two random variables X and Y is given by*

$$\begin{aligned}
\sigma_{XY} &= cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \\
&= \sum_{x \in S_X} \sum_{y \in S_Y} (x - \mu_X)(y - \mu_Y) \Pr(X = x, Y = y) \quad \textit{for discrete } X \textit{ and } Y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) p(x, y) dx dy \quad \textit{for continuous } X \textit{ and } Y
\end{aligned}$$

De&nition 37 *The correlation between two random variables X and Y is given by*

$$\rho_{XY} = corr(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Notice that the correlation coefficient, ρ_{XY}, is just a scaled version of the covariance.

To see how covariance measures the direction of linear association, consider the probability scatterplot in &gure xxx.

(insert &gure here)

In the plot the random variables X and Y are distributed such that $\mu_X = \mu_Y = 0$. The plot is separated into quadrants. In the &rst quandrant, the realized values satisfy $x < \mu_X, y > \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) < 0$. In the second quadrant, the values satisfy $x > \mu_X$ and $y > \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) > 0$. In the third quadrant, the values satisfy $x > \mu_X$ but $y < \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) < 0$. Finally, in the fourth quandrant, $x < \mu_X$ and $y < \mu_Y$ so that the product $(x - \mu_X)(y - \mu_Y) > 0$. Covariance is then a probability weighted average all of the product terms in the four quadrants. For the example data, this weighted average turns out to be positive.

Example 38 *For the data in Table 2, we have*

$$\begin{aligned}
\sigma_{XY} &= cov(X,Y) = (0-3/2)(0-1/2)\cdot 1/8 + (0-3/2)(1-1/2)\cdot 0 + \cdots + (3-3/2)(1-1/2)\cdot 1/8 \\
\rho_{XY} &= corr(X,Y) = \frac{1/4}{\sqrt{(3/4)\cdot(1/2)}} = 0.577
\end{aligned}$$

2.5.1 Properties of Covariance and Correlation

Let X and Y be random variables and let a and b be constants. Some important properties of $cov(X,Y)$ are

1. $cov(X,X) = var(X)$
2. $cov(X,Y) = cov(Y,X)$
3. $cov(aX,bY) = a \cdot b \cdot cov(X,Y)$
4. If X and Y are independent then $cov(X,Y) = 0$ (no association $\Longrightarrow$ no linear association). However, if $cov(X,Y) = 0$ then X and Y are not necessarily independent (no linear association $\not\Rightarrow$ no association).
5. If X and Y are jointly normally distributed and $cov(X,Y) = 0$, then X and Y are independent.

The third property above shows that the value of $cov(X,Y)$ depends on the scaling of the random variables X and Y. By simply changing the scale of X or Y we can make $cov(X,Y)$ equal to any value that we want. Consequently, the numerical value of $cov(X,Y)$ is not informative about the strength of the linear association between X and Y. However, the sign of $cov(X,Y)$ is informative about the direction of linear association between X and Y. The fourth property should be intuitive. Independence between the random variables X and Y means that there is no relationship, linear or nonlinear, between X and Y. However, the lack of a linear relationship between X and Y does not preclude a nonlinear relationship. The last result illustrates an important property of the normal distribution: lack of covariance implies independence.

Some important properties of $corr(X,Y)$ are

1. $-1 \leq \rho_{XY} \leq 1$.
2. If $\rho_{XY} = 1$ then X and Y are perfectly positively linearly related. That is, $Y = aX + b$ where $a > 0$.
3. If $\rho_{XY} = -1$ then X and Y are perfectly negatively linearly related. That is, $Y = aX + b$ where $a < 0$.
4. If $\rho_{XY} = 0$ then X and Y are not linearly related but may be nonlinearly related.
5. $corr(aX, bY) = corr(X,Y)$ if $a > 0$ and $b > 0$; $corr(X,Y) = -corr(X,Y)$ if $a > 0, b < 0$ or $a < 0, b > 0$.

(Stuff to add: bivariate normal distribution)

2.5.2 Expectation and variance of the sum of two random variables

Let X and Y be two random variables with well de&ned means, variances and covariance and let a and b be constants. Then the following results hold.

1. $E[aX + bY] = aE[X] + bE[Y] = a\mu_X + b\mu_Y$
2. $var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2 \cdot a \cdot b \cdot cov(X,Y) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2 \cdot a \cdot b \cdot \sigma_{XY}$

The &rst result states that the expected value of a linear combination of two random variables is equal to a linear combination of the expected values of the random variables. This result indicates that the expectation operator is a linear operator. In other words, expectation is additive. The second result states that variance of a linear combination of random variables is not a linear combination of the variances of the random variables. In particular, notice that covariance comes up as a term when computing the variance of the sum of two (not independent) random variables.

Hence, the variance operator is not, in general, a linear operator. That is, variance, in general, is not additive.

It is worthwhile to go through the proofs of these results, at least for the case of discrete random variables. Let X and Y be discrete random variables. Then,

$$\begin{aligned}
E[aX+bY] &= \sum_{x\in S_X}\sum_{y\in S_y}(ax+by)\Pr(X=x,Y=y) \\
&= \sum_{x\in S_X}\sum_{y\in S_y}ax\Pr(X=x,Y=y)+\sum_{x\in S_X}\sum_{y\in S_y}bx\Pr(X=x,Y=y) \\
&= a\sum_{x\in S_X}x\sum_{y\in S_y}\Pr(X=x,Y=y)+b\sum_{y\in S_y}y\sum_{x\in S_X}\Pr(X=x,Y=y) \\
&= a\sum_{x\in S_X}x\Pr(X=x)+b\sum_{y\in S_y}y\Pr(Y=y) \\
&= aE[X]+bE[Y]=a\mu_X+b\mu_Y.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
var(aX+bY) &= E[(aX+bY-E[aX+bY])^2] \\
&= E[(aX+bY-a\mu_X-b\mu_Y)^2] \\
&= E[(a(X-\mu_X)+b(Y-\mu_Y))^2] \\
&= a^2E[(X-\mu_X)^2]+b^2E[(Y-\mu_Y)^2]+2\cdot a\cdot b\cdot E[(X-\mu_X)(Y-\mu_Y)] \\
&= a^2var(X)+b^2var(Y)+2\cdot a\cdot b\cdot cov(X,Y).
\end{aligned}$$

2.5.3 Linear Combination of two Normal random variables

The following proposition gives an important result concerning a linear combination of normal random variables.

Proposition 39 *Let $X \sim N(\mu_X,\sigma_X^2)$, $Y \sim N(\mu_Y,\sigma_Y^2)$, $\sigma_{XY}=cov(X,Y)$ and a and b be constants. Define the new random variable Z as*

$$Z = aX+bY.$$

Then

$$Z \sim N(\mu_Z,\sigma_Z^2)$$

where

$$\begin{aligned}
\mu_Z &= a\mu_X+b\mu_Y \\
\sigma_Z^2 &= a^2\sigma_X^2+b^2\sigma_Y^2+2ab\sigma_{XY}
\end{aligned}$$

This important result states that a linear combination of two normally distributed random variables is itself a normally distributed random variable. The proof of the result relies on the change of variables theorem from calculus and is omitted. Not all random variables have the property that their distributions are closed under addition.

3 Multivariate Distributions

The results for bivariate distributions generalize to the case of more than two random variables. The details of the generalizations are not important for our purposes. However, the following results will be used repeatedly.

3.1 Linear Combinations of N Random Variables

Let $X_1, X_2, \ldots, X_N$ denote a collection of N random variables with means μ_i,variances σ_i^2 and covariances σ_{ij}. De&ne the new random variable Z as a linear combination

$$Z = a_1 X_1 + a_2 X_2 + \cdots + a_N X_N$$

where $a_1, a_2, \ldots, a_N$ are constants. Then the following results hold

$$\begin{aligned}
\mu_Z &= E[Z] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_N E[X_N] \\
&= \sum_{i=1}^{N} a_i E[X_i] = \sum_{i=1}^{N} a_i \mu_i.
\end{aligned}$$

$$\begin{aligned}
\sigma_Z^2 &= var(Z) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_N^2\sigma_N^2 \\
&\quad +2a_1a_2\sigma_{12} + 2a_1a_3\sigma_{13} + \cdots + a_1a_N\sigma_{1N} \\
&\quad +2a_2a_3\sigma_{23} + 2a_2a_4\sigma_{24} + \cdots + a_2a_N\sigma_{2N} \\
&\quad + \cdots + \\
&\quad +2a_{N-1}a_N\sigma_{(N-1)N}
\end{aligned}$$

In addition, if all of the X_i are normally distributed then Z is normally distributed with mean μ_Z and variance σ_Z^2 as described above.

3.1.1 Application: Distribution of Continuously Compounded Returns

Let R_t denote the continuously compounded monthly return on an asset at time t. Assume that R_t ˜ $iid\ N(\mu, \sigma^2)$. The annual continuously compounded return is equal the sum of twelve monthly continuously compounded returns. That is,

$$R_t(12) = \sum_{j=0}^{11} R_{t-j}.$$

Since each monthly return is normally distributed, the annual return is also normally distributed. In addition,

$$
\begin{aligned}
E[R_t(12)] &= E\left[\sum_{j=0}^{11} R_{t-j}\right] \\
&= \sum_{j=0}^{11} E[R_{t-j}] \text{ (by linearity of expectation)} \\
&= \sum_{j=0}^{11} \mu \quad \text{(by identical distributions)} \\
&= 12 \cdot \mu,
\end{aligned}
$$

so that the expected annual return is equal to 12 times the expected monthly return. Furthermore,

$$
\begin{aligned}
var(R_t(12)) &= var\left(\sum_{j=0}^{11} R_{t-j}\right) \\
&= \sum_{j=0}^{11} var(R_{t-j}) \quad \text{(by independence)} \\
&= \sum_{j=0}^{11} \sigma^2 \quad \text{(by identical distributions)} \\
&= 12 \cdot \sigma^2,
\end{aligned}
$$

so that the annual variance is also equal to 12 times the monthly variance[2]. For the annual standard deviation, we have

$$SD(R_t(12)) = \sqrt{12}\sigma.$$

4 Further Reading

Excellent intermediate level treatments of probability theory using calculus are given in DeGroot (1986), Hoel, Port and Stone (1971) and Hoag and Craig (19xx). Intermediate treatments with an emphasis towards applications in &nance include Ross (1999) and Watsom and Parramore (1998). Intermediate textbooks with an emphasis on econometrics include Amemiya (1994), Goldberger (1991), Ramanathan (1995). Advanced treatments of probability theory applied to &nance are given in Neftci (1996). Everything you ever wanted to know about probability distributions is given Johnson and Kotz (19xx).

[2] This result often causes some confusion. It is easy to make the mistake and say that the annual variance is $(12)^2 = 144$ time the monthly variance. This result would occur if $R_A = 12R_t$, so that $var(R_A) = (12)^2 var(R_t) = 144 var(R_t)$.

5 Problems

Let $W, X, Y,$ and Z be random variables describing next year's annual return on Weyerhauser, Xerox, Yahoo! and Zymogenetics stock. The table below gives discrete probability distributions for these random variables based on the state of the economy:

State of Economy	W	$p(w)$	X	$p(x)$	Y	$p(y)$	Z	$p(z)$
Depression	-0.3	0.05	-0.5	0.05	-0.5	0.15	-0.8	0.05
Recession	0.0	0.2	-0.2	0.1	-0.2	0.5	0.0	0.2
Normal	0.1	0.5	0	0.2	0	0.2	0.1	0.5
Mild Boom	0.2	0.2	0.2	0.5	0.2	0.1	0.2	0.2
Major Boom	0.5	0.05	0.5	0.15	0.5	0.05	1	0.05

- Plot the distributions for each random variable (make a bar chart). Comment on any differences or similarities between the distributions.
- For each random variable, compute the expected value, variance, standard deviation, skewness, kurtosis and brie! y comment.

Suppose X is a normally distributed random variable with mean 10 and variance 24.

- Find $\Pr(X > 14)$
- Find $\Pr(8 < X < 20)$
- Find the probability that X takes a value that is at least 6 away from its mean.
- Suppose y is a constant de&ned such that $\Pr(X > y) = 0.10$. What is the value of y?
- Determine the $1\%, 5\%, 10\%, 25\%$ and 50% quantiles of the distribution of X.

Let X denote the monthly return on Microsoft stock and let Y denote the monthly return on Starbucks stock. Suppose $X\tilde{}N(0.05, (0.10)^2)$ and $Y\tilde{}N(0.025, (0.05)^2)$.

- Plot the normal curves for X and Y
- Comment on the risk-return trade-offs for the two stocks

Let R denote the monthly return on Microsoft stock and let W_0 denote initial wealth to be invested in Microsoft stock over the next month. Assume that $R\tilde{}N(0.07, (0.12)^2)$ and that $W_0 = \$25,000$.

- What is the distribution of end of month wealth $W_1 = W_0(1 + R)$?
- What is the probability that end of month wealth is less than \$20,000?
- What is the Value-at-Risk (VaR) on the investment in Microsoft stock over the next month with 5% probability?

References

[1] Amemiya, T. (1994). *Introduction to Statistics and Econometrics*. Harvard University Press, Cambridge, MA.

[2] Goldberger, A.S. (1991). *A Course in Econometrics*. Harvard University Press, Cambridge, MA.

[3] Hoel, P.G., Port, S.C. and Stone, C.J. (1971). *Introduction to Probability Theory*. Houghton Mifflin, Boston, MA.

[4] Johnson, x, and Kotz, x. *Probability Distributions*, Wiley.

[5] Neftci, S.N. (1996). *An Introduction to the Mathematics of Financial Derivatives*. Academic Press, San Diego, CA.

[6] Ross, S. (1999). *An Introduction to Mathematical Finance: Options and Other Topics*. Cambridge University Press, Cambridge, UK.

[7] Watsham, T.J., and Parramore, K. (1998). *Quantitative Methods in Finance*. International Thompson Business Press, London, UK.

Introduction to Financial Econometrics
Chapter 3 The Constant Expected Return Model

Eric Zivot
Department of Economics
University of Washington

January 6, 2000
This version: January 23, 2001

1 The Constant Expected Return Model of Asset Returns

1.1 Assumptions

Let R_{it} denote the continuously compounded return on an asset i at time t. We make the following assumptions regarding the probability distribution of R_{it} for $i = 1, \ldots, N$ assets over the time horizon $t = 1, \ldots, T$.

1. Normality of returns: $R_{it} \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
2. Constant variances and covariances: $cov(R_{it}, R_{jt}) = \sigma_{ij}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
3. No serial correlation across assets over time: $cov(R_{it}, R_{js}) = 0$ for $t \neq s$ and $i, j = 1, \ldots, N$.

Assumption 1 states that in every time period asset returns are normally distributed and that the mean and the variance of each asset return is constant over time. In particular, we have for each asset i

$$\begin{aligned} E[R_{it}] &= \mu_i \text{ for all values of } t \\ var(R_{it}) &= \sigma_i^2 \text{ for all values of } t \end{aligned}$$

The second assumption states that the contemporaneous covariances between assets are constant over time. Given assumption 1, assumption 2 implies that the contemporaneous correlations between assets are constant over time as well. That is, for all assets

$$corr(R_{it}, R_{jt}) = \rho_{ij} \text{ for all values of } t.$$

The third assumption stipulates that all of the asset returns are uncorrelated over time[1]. In particular, for a given asset i the returns on the asset are *serially uncorrelated* which implies that

$$corr(R_{it}, R_{is}) = cov(R_{it}, R_{is}) = 0 \text{ for all } t \neq s.$$

Additionally, the returns on all possible pairs of assets i and j are serially uncorrelated which implies that

$$corr(R_{it}, R_{js}) = cov(R_{it}, R_{js}) = 0 \text{ for all } i \neq j \text{ and } t \neq s.$$

Assumptions 1-3 indicate that all asset returns at a given point in time are jointly (multivariate) normally distributed and that this joint distribution stays constant over time. Clearly these are very strong assumptions. However, they allow us to development a straightforward probabilistic model for asset returns as well as statistical tools for estimating the parameters of the model and testing hypotheses about the parameter values and assumptions.

1.2 Constant Expected Return Model Representation

A convenient mathematical representation or *model* of asset returns can be given based on assumptions 1-3. This is the *constant expected return* (CER) model. For assets $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$ the CER model is represented as

$$\begin{aligned} R_{it} &= \mu_i + \varepsilon_{it} & (1) \\ \varepsilon_{it} &\sim \; i.i.d.\ N(0, \sigma_i^2) & \\ cov(\varepsilon_{it}, \varepsilon_{jt}) &= \sigma_{ij} & (2) \end{aligned}$$

where μ_i is a constant and we assume that ε_{it} is independent of ε_{js} for all time periods $t \neq s$. The notation $\varepsilon_{it} \sim i.i.d.\ N(0, \sigma_i^2)$ stipulates that the random variable ε_{it} is serially independent and identically distributed as a normal random variable with mean zero and variance σ_i^2. In particular, note that, $E[\varepsilon_{it}] = 0$, $var(\varepsilon_{it}) = \sigma_i^2$ and $cov(\varepsilon_{it}, \varepsilon_{js}) = 0$ for $i \neq j$ and $t \neq s$.

Using the basic properties of expectation, variance and covariance discussed in chapter 2, we can derive the following properties of returns. For expected returns we have

$$E[R_{it}] = E[\mu_i + \varepsilon_{it}] = \mu_i + E[\varepsilon_{it}] = \mu_i,$$

[1]Since all assets are assumed to be normally distributed (assumption 1), uncorrelatedness implies the stronger condition of independence.

since μ_i is constant and $E[\varepsilon_{it}] = 0$. Regarding the variance of returns, we have

$$var(R_{it}) = var(\mu_i + \varepsilon_{it}) = var(\varepsilon_{it}) = \sigma_i^2$$

which uses the fact that the variance of a constant (μ_i) is zero. For covariances of returns, we have

$$cov(R_{it}, R_{jt}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{jt}) = cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}$$

and

$$cov(R_{it}, R_{js}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{js}) = cov(\varepsilon_{it}, \varepsilon_{js}) = 0,\ t \neq s,$$

which use the fact that adding constants to two random variables does not affect the covariance between them. Given that covariances and variances of returns are constant over time gives the result that correlations between returns over time are also constant:

$$\begin{aligned} corr(R_{it}, R_{jt}) &= \frac{cov(R_{it}, R_{jt})}{\sqrt{var(R_{it})var(R_{jt})}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij}, \\ corr(R_{it}, R_{js}) &= \frac{cov(R_{it}, R_{js})}{\sqrt{var(R_{it})var(R_{js})}} = \frac{0}{\sigma_i \sigma_j} = 0,\ i \neq j, t \neq s. \end{aligned}$$

Finally, since the random variable ε_{it} is independent and identically distributed (*i.i.d.*) normal the asset return R_{it} will also be *i.i.d.* normal:

$$R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2).$$

Hence, the CER model (1) for R_{it} is equivalent to the model implied by assumptions 1-3.

1.3 Interpretation of the CER Model

The CER model has a very simple form and is identical to the *measurement error model* in the statistics literature. In words, the model states that each asset return is equal to a constant μ_i (the expected return) plus a normally distributed random variable ε_{it} with mean zero and constant variance. The random variable ε_{it} can be interpreted as representing the unexpected news concerning the value of the asset that arrives between times $t-1$ and time t. To see this, note that using (1) we can write ε_{it} as

$$\begin{aligned} \varepsilon_{it} &= R_{it} - \mu_i \\ &= R_{it} - E[R_{it}] \end{aligned}$$

so that ε_{it} is de&ned to be the deviation of the random return from its expected value. If the news is good, then the realized value of ε_{it} is positive and the observed return is above its expected value μ_i. If the news is bad, then ε_{jt} is negative and the observed return is less than expected. The assumption that $E[\varepsilon_{it}] = 0$ means that news, on average, is neutral; neither good nor bad. The assumption that $var(\varepsilon_{it}) = \sigma_i^2$ can be interpreted as saying that volatility of news arrival is constant over time. The random news variable affecting asset i, e_{it}, is allowed to be contemporaneously correlated with the random news variable affecting asset j, ε_{jt}, to capture the idea that news about one asset may spill over and affect another asset. For example, let asset i be Microsoft and asset j be Apple Computer. Then one interpretation of news in this context is general news about the computer industry and technology. Good news should lead to positive values of ε_{it} and ε_{jt}. Hence these variables will be positively correlated.

The CER model with continuously compounded returns has the following nice property with respect to the interpretation of ε_{it} as news. Consider the default case where R_{it} is interpreted as the continuously compounded monthly return. Since multiperiod continuously compounded returns are additive we can interpret, for example, R_{it} as the sum of 30 daily continuously compounded returns[2]:

$$R_{it} = \sum_{k=0}^{29} R_{it-k}^d$$

where R_{it}^d denotes the continuously compounded daily return on asset i. If we assume that daily returns are described by the CER model then

$$\begin{aligned}
R_{it}^d &= \mu_i^d + \varepsilon_{it}^d, \\
\varepsilon_{it}^d &\sim \; i.i.d\ N(0, (\sigma_i^d)^2), \\
cov(\varepsilon_{it}^d, \varepsilon_{jt}^d) &= \sigma_{ij}^d, \\
cov(\varepsilon_{it}^d, \varepsilon_{js}^d) &= 0,\ i \neq j, t \neq s
\end{aligned}$$

and the monthly return may then be expressed as

$$\begin{aligned}
R_{it} &= \sum_{k=0}^{29} (\mu_i^d + \varepsilon_{it-k}^d) \\
&= 30 \cdot \mu_i^d + \sum_{k=0}^{29} \varepsilon_{it-k}^d \\
&= \mu_i + \varepsilon_{it},
\end{aligned}$$

where

$$\begin{aligned}
\mu_i &= 30 \cdot \mu_i^d, \\
\varepsilon_{it} &= \sum_{k=0}^{29} \varepsilon_{it-k}^d.
\end{aligned}$$

[2]For simplicity of exposition, we will ignore the fact that some assets do not trade over the weekend.

Hence, the monthly expected return, μ_i, is simply 30 times the daily expected return. The interpretation of ε_{it} in the CER model when returns are continuously compounded is the accumulation of news between months $t-1$ and t. Notice that

$$\begin{aligned} var(R_{it}) &= var\left(\sum_{k=0}^{29}(\mu_i^d + \varepsilon_{it-k}^d)\right) \\ &= \sum_{k=0}^{29} var(\varepsilon_{it-k}^d) \\ &= \sum_{k=0}^{29}\left(\sigma_i^d\right)^2 \\ &= 30 \cdot \left(\sigma_i^d\right)^2 \end{aligned}$$

and

$$\begin{aligned} cov(R_{it}, R_{jt}) &= cov\left(\sum_{k=0}^{29}\varepsilon_{it-k}^d, \sum_{k=0}^{29}\varepsilon_{jt-k}^d\right) \\ &= \sum_{k=0}^{29} cov(\varepsilon_{it-k}^d, \varepsilon_{jt-k}^d) \\ &= \sum_{k=0}^{29}\sigma_{ij}^d \\ &= 30 \cdot \sigma_{ij}^d, \end{aligned}$$

so that the monthly variance, σ_i^2, is equal to 30 times the daily variance and the monthly covariance, σ_{ij}, is equal to 30 times the daily covariance.

1.4 The CER Model of Asset Returns and the Random Walk Model of Asset Prices

The CER model of asset returns (1) gives rise to the so-called *random walk* (RW) model of the logarithm of asset prices. To see this, recall that the continuously compounded return, R_{it}, is de&ned from asset prices via

$$\ln\left(\frac{P_{it}}{P_{it-1}}\right) = R_{it}.$$

Since the log of the ratio of prices is equal to the difference in the logs of prices we may rewrite the above as

$$\ln(P_{it}) - \ln(P_{it-1}) = R_{it}.$$

Letting $p_{it} = \ln(P_{it})$ and using the representation of R_{it} in the CER model (1), we may further rewrite the above as

$$p_{it} - p_{it-1} = \mu_i + \varepsilon_{it}. \tag{3}$$

The representation in (3) is know as the RW model for the log of asset prices.

In the RW model, μ_i represents the expected change in the log of asset prices (continuously compounded return) between months $t-1$ and t and ε_{it} represents the unexpected change in prices. That is,

$$\begin{aligned} E[p_{it} - p_{it-1}] &= E[R_{it}] = \mu_i, \\ \varepsilon_{it} &= p_{it} - p_{it-1} - E[p_{it} - p_{it-1}]. \end{aligned}$$

Further, in the RW model, the unexpected changes in asset prices, ε_{it}, are uncorrelated over time ($cov(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $t \neq s$) so that future changes in asset prices cannot be predicted from past changes in asset prices[3].

The RW model gives the following interpretation for the evolution of asset prices. Let p_{i0} denote the initial log price of asset i. The RW model says that the price at time $t = 1$ is

$$p_{i1} = p_{i0} + \mu_i + \varepsilon_{i1}$$

where ε_{i1} is the value of random news that arrives between times 0 and 1. Notice that at time $t = 0$ the expected price at time $t = 1$ is

$$E[p_{i1}] = p_{i0} + \mu_i + E[\varepsilon_{i1}] = p_{i0} + \mu_i$$

which is the initial price plus the expected return between time 0 and 1. Similarly, the price at time $t = 2$ is

$$\begin{aligned} p_{i2} &= p_{i1} + \mu_i + \varepsilon_{i2} \\ &= p_{i0} + \mu_i + \mu_i + \varepsilon_{i1} + \varepsilon_{i2} \\ &= p_{i0} + 2 \cdot \mu_i + \sum_{t=1}^{2} \varepsilon_{it} \end{aligned}$$

which is equal to the initial price, p_{i0}, plus the two period expected return, $2 \cdot \mu_i$, plus the accumulated random news over the two periods, $\sum_{t=1}^{2} \varepsilon_{it}$. By recursive substitution, the price at time $t = T$ is

$$p_{iT} = p_{i0} + T \cdot \mu_i + \sum_{t=1}^{T} \varepsilon_{it}.$$

At time $t = 0$ the expected price at time $t = T$ is

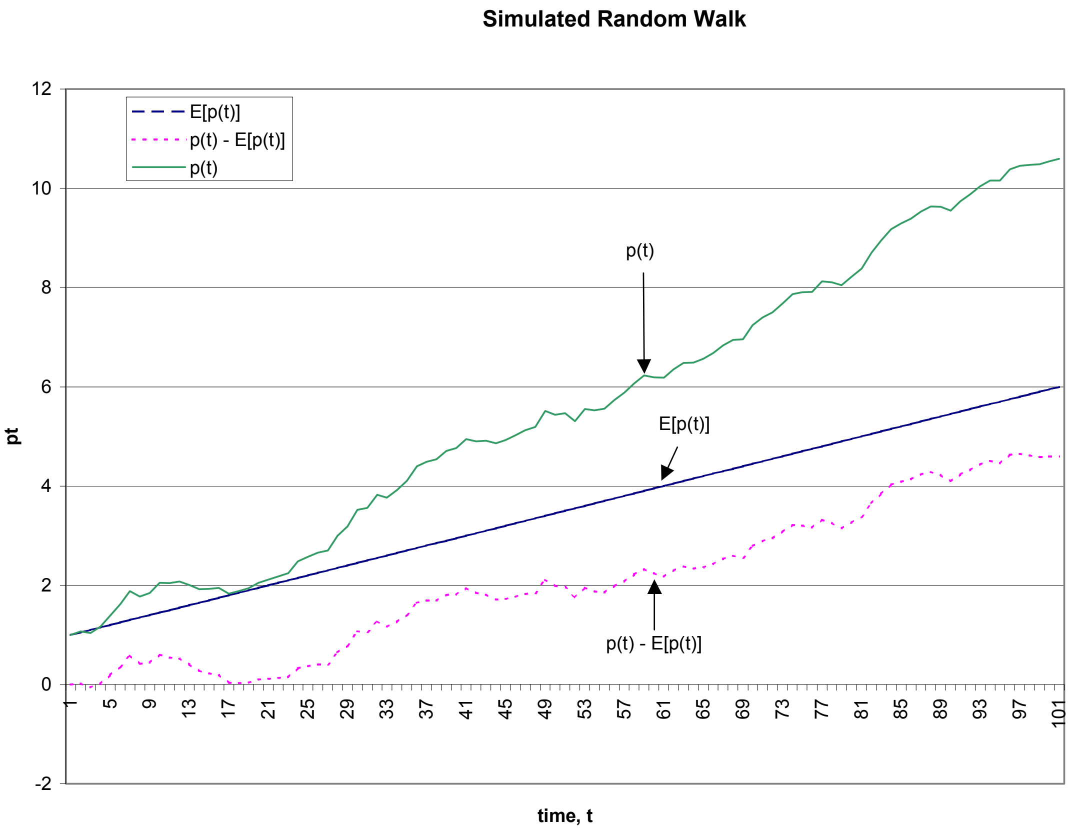
$$E[p_{iT}] = p_{i0} + T \cdot \mu_i$$

The actual price, p_{iT}, deviates from the expected price by the accumulated random news

$$p_{iT} - E[p_{iT}] = \sum_{t=1}^{T} \varepsilon_{it}.$$

Figure xxx illustrates the random walk model of asset prices.

[3] The notion that future changes in asset prices cannot be predicted from past changes in asset prices is often referred to as the weak form of the efficient markets hypothesis.



The term *random walk* was originally used to describe the unpredictable movements of a drunken sailor staggering down the street. The sailor starts at an initial position, p_0, outside the bar. The sailor generally moves in the direction described by μ but randomly deviates from this direction after each step t by an amount equal to ε_t. After T steps the sailor ends up at position $p_T = p_0 + \mu \cdot T + \sum_{t=1}^{T} \varepsilon_t$.

2 Monte Carlo Simulation of the CER Model

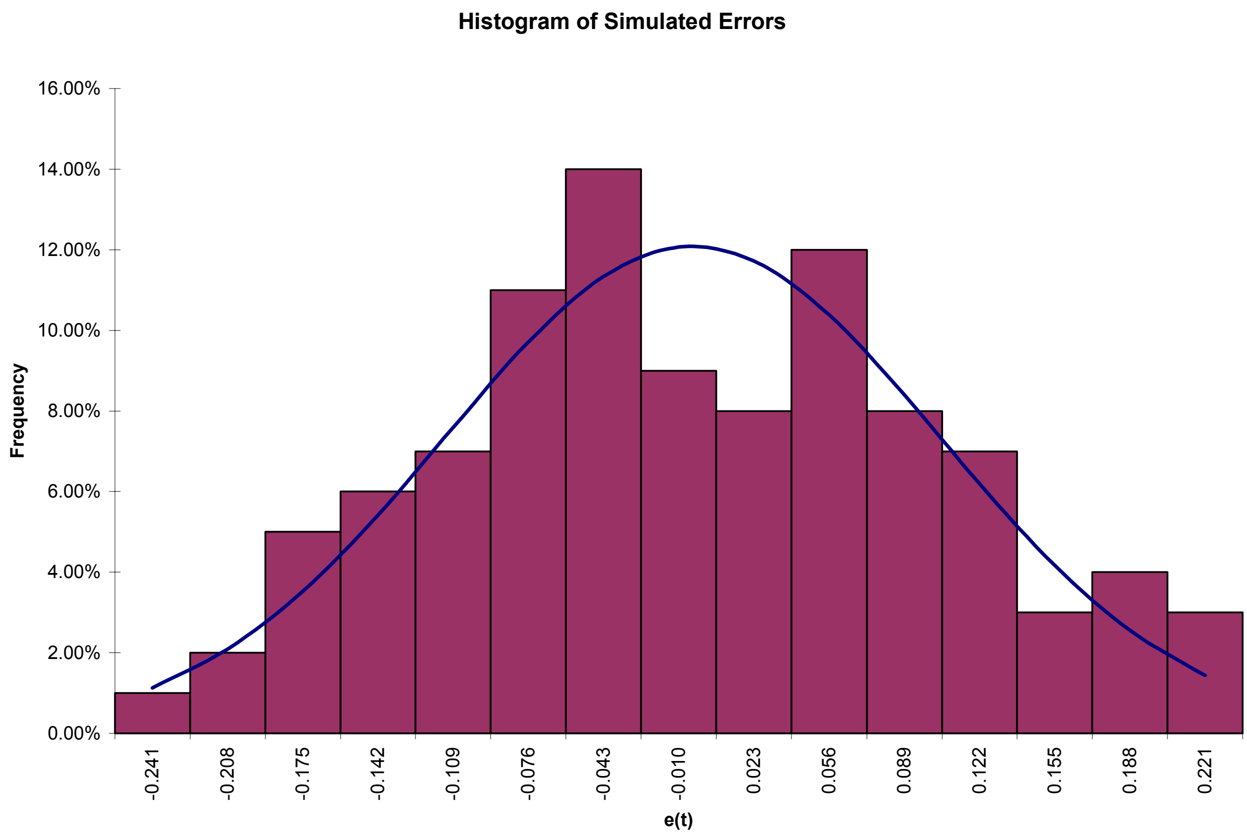
A good way to understand the probabilistic behavior of a model is to use computer simulation methods to create pseudo data from the model. The process of creating such pseudo data is often called *Monte Carlo simulation*[4]. To illustrate the use of Monte Carlo simulation, consider the problem of creating pseudo return data from the CER model (1) for one asset. In order to simulate pseudo return data, values for the model parameters μ and σ must be selected. To mimic the monthly return data on Microsoft, the values $\mu = 0.05$ and $\sigma = 0.10$ are used. Also, the number N of

[4]Monte Carlo referrs to the fameous city in Monaco where gambling is legal.

simulated data points must be determined. Here, $N = 100$. Hence, the model to be simulated is

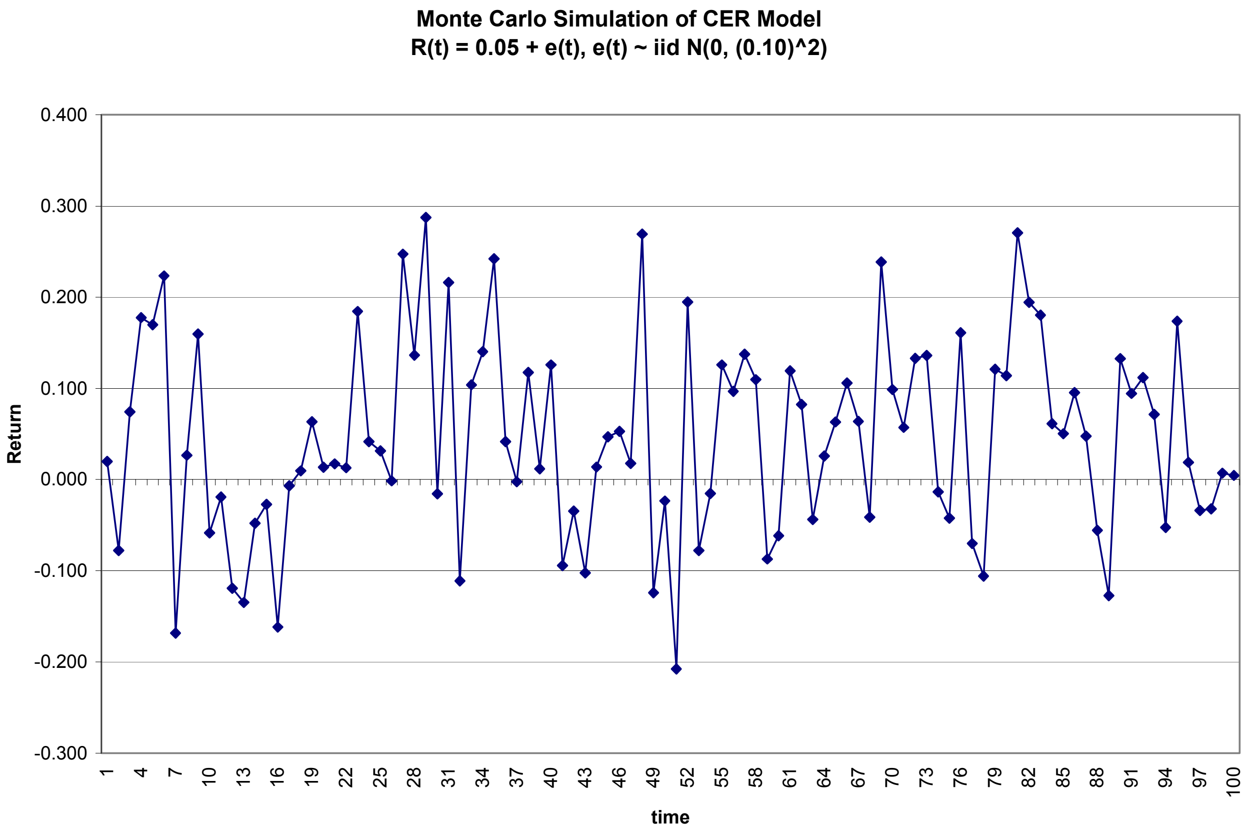
$$R_t = 0.05 + \varepsilon_t, \ t = 1, \ldots, 100$$
$$\varepsilon_t \text{ ~} iid \ N(0, (0.10)^2)$$

The key to simulating data from the above model is to simulate $N = 100$ observations of the random news variable ε_t ~$iid \ N(0, (0.10)^2)$. Computer algorithms exist which can easily create such observations. Let $\{\varepsilon_1, \ldots, \varepsilon_{100}\}$ denote the 100 simulated values of ε_t. The histogram of these values are given in &gure xxx below



The sample average of the simulated errors is $\frac{1}{100}\sum_{t=1}^{100}\varepsilon_t = -0.004$ and the sample standard deviation is $\sqrt{\frac{1}{99}\sum_{t=1}^{100}(\varepsilon_t - (-0.004))^2} = 0.109$. These values are very close to the population values $E[\varepsilon_t] = 0$ and $SD(\varepsilon_t) = 0.10$, respectively.

Once the simulated values of ε_t have been created, the simulated values of R_t are constructed as $R_t = 0.05 + \varepsilon_t$, $t = 1, \ldots, 100$. A time plot of the simulated values of R_t is given in &gure xxx below



The simulated return data fluctuates randomly about the expected return value $E[R_t] = \mu = 0.05$. The typical size of the fluctuation is approximately equal to $SE(\varepsilon_t) = 0.10$. Notice that the simulated return data looks remarkably like the actual return data of Microsoft.

Monte Carlo simulation of a model can be used as a first pass reality check of the model. If simulated data from the model does not look like the data that the model is supposed to describe then serious doubt is cast on the model. However, if simulated data looks reasonably close to the data that the model is suppose to describe then confidence is instilled on the model.

3 Estimating the CER Model

3.1 The Random Sampling Environment

The CER model of asset returns gives us a rigorous way of interpreting the time series behavior of asset returns. At the beginning of every month t, R_{it} is a random variable representing the return to be realized at the end of the month. The CER model states that $R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2)$. Our best guess for the return at the end of the month is $E[R_{it}] = \mu_i$, our measure of uncertainty about our best guess is captured by $\sigma_i = \sqrt{var(R_{it})}$ and our measure of the direction of linear association between R_{it} and R_{jt} is $\sigma_{ij} = cov(R_{it}, R_{jt})$. The CER model assumes that the economic environment is constant over time so that the normal distribution characterizing monthly returns is the same every month.

Our life would be very easy if we knew the exact values of μ_i, σ_i^2 and σ_{ij}, the parameters of the CER model. In actuality, however, we do not know these values with certainty. A key task in &nancial econometrics is estimating the values of μ_i, σ_i^2 and σ_{ij} from a history of observed data.

Suppose we observe monthly returns on N different assets over the horizon $t = 1, \ldots, T$. Let $r_{i1}, \ldots, r_{iT}$ denote the observed history of T monthly returns on asset i for $i = 1, \ldots, N$. It is assumed that the observed returns are realizations of the random variables $R_{i1}, \ldots, R_{iT}$, where R_{it} is described by the CER model (1). We call $R_{i1}, \ldots, R_{iT}$ a random sample from the CER model (1) and we call $r_{i1}, \ldots, r_{iT}$ the realized values from the random sample. In this case, we can use the observed returns to estimate the unknown parameters of the CER model

3.2 Estimation Theory

Before we describe the estimation of the CER model, it is useful to summarize some concepts in estimation theory. Let θ denote some characteristic of the CER model (1) we are interested in estimating. For example, if we are interested in the expected return then $\theta = \mu_i$; if we are interested in the variance of returns then $\theta = \sigma_i^2$. The goal is to estimate θ based on the observed data $r_{i1}, \ldots, r_{iT}$.

De&nition 1 *An estimator of θ is a rule or algorithm for forming an estimate for θ.*

De&nition 2 *An estimate of θ is simply the value of an estimator based on the observed data.*

To establish some notation, let $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ denote an estimator of θ treated as a function of the random variables $R_{i1}, \ldots, R_{iT}$. Clearly, $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ is a random variable. Let $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ denote an estimate of θ based on the realized values $r_{i1}, \ldots, r_{iT}$. $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ is simply an number. We will often use $\hat{\theta}$ as shorthand notation to represent either an estimator of θ or an estimate of θ. The context will determine how to interpret $\hat{\theta}$.

3.2.1 Properties of Estimators

Consider $\hat{\theta} = \hat{\theta}(R_{i1}, \ldots, R_{iT})$ as a random variable. In general, the pdf of $\hat{\theta}$, $p(\hat{\theta})$, depends on the pdf's of the random variables $R_{i1}, \ldots, R_{iT}$. The exact form of $p(\hat{\theta})$ may

Introduction to Financial Econometrics
Chapter 4 Introduction to Portfolio Theory

Eric Zivot
Department of Economics
University of Washington

January 26, 2000
This version: February 20, 2001

1 Introduction to Portfolio Theory

Consider the following investment problem. We can invest in two non-dividend paying stocks A and B over the next month. Let R_A denote monthly return on stock A and R_B denote the monthly return on stock B. These returns are to be treated as random variables since the returns will not be realized until the end of the month. We assume that the returns R_A and R_B are jointly normally distributed and that we have the following information about the means, variances and covariances of the probability distribution of the two returns:

$$\begin{aligned}
\mu_A &= E[R_A],\ \sigma_A^2 = Var(R_A), \\
\mu_B &= E[R_B],\ \sigma_B^2 = Var(R_B), \\
\sigma_{AB} &= Cov(R_A, R_B).
\end{aligned}$$

We assume that these values are taken as given. We might wonder where such values come from. One possibility is that they are estimated from historical return data for the two stocks. Another possibility is that they are subjective guesses.

The expected returns, μ_A and μ_B, are our best guesses for the monthly returns on each of the stocks. However, since the investments are random we must recognize that the realized returns may be different from our expectations. The variances, σ_A^2 and σ_B^2, provide measures of the uncertainty associated with these monthly returns. We can also think of the variances as measuring the risk associated with the investments. Assets that have returns with high variability (or volatility) are often thought to be risky and assets with low return volatility are often thought to be safe. The covariance σ_{AB} gives us information about the *direction* of any linear dependence between returns. If $\sigma_{AB} > 0$ then the returns on assets A and B tend to move in the same direction; if $\sigma_{AB} < 0$ the returns tend to move in opposite directions; if $\sigma_{AB} = 0$ then the returns tend to move independently. The strength of the dependence between the returns is measured by the correlation coefficient $\rho_{AB} = \frac{\sigma_{AB}}{\sigma_A \sigma_B}$. If ρ_{AB} is close to one in absolute value then returns mimic each other extremely closely whereas if ρ_{AB} is close to zero then the returns may show very little relationship.

The portfolio problem is set-up as follows. We have a given amount of wealth and it is assumed that we will exhaust all of our wealth between investments in the two stocks. The investor's problem is to decide how much wealth to put in asset A and how much to put in asset B. Let x_A denote the share of wealth invested in stock A and x_B denote the share of wealth invested in stock B. Since all wealth is put into the two investments it follows that $x_A + x_B = 1$. (Aside: What does it mean for x_A or x_B to be negative numbers?) The investor must choose the values of x_A and x_B.

Our investment in the two stocks forms a *portfolio* and the shares x_A and x_B are referred to as *portfolio shares* or weights. The return on the portfolio over the next month is a random variable and is given by

$$R_p = x_A R_A + x_B R_B, \tag{1}$$

which is just a simple linear combination or weighted average of the random return variables R_A and R_B. Since R_A and R_B are assumed to be normally distributed, R_p is also normally distributed.

1.1 Portfolio expected return and variance

The return on a portfolio is a random variable and has a probability distribution that depends on the distributions of the assets in the portfolio. However, we can easily deduce some of the properties of this distribution by using the following results concerning linear combinations of random variables:

$$\begin{aligned} \mu_p &= E[R_p] = x_A \mu_A + x_B \mu_B && (2) \\ \sigma_p^2 &= var(R_p) = x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2 x_A x_B \sigma_{AB} && (3) \end{aligned}$$

These results are so important to portfolio theory that it is worthwhile to go through the derivations. For the &rst result (2), we have

$$E[R_p] = E[x_A R_A + x_B R_B] = x_A E[R_A] + x_B E[R_B] = x_A \mu_A + x_B \mu_B$$

by the linearity of the expectation operator. For the second result (3), we have

$$\begin{aligned} var(R_p) &= var(x_A R_A + x_B R_B) = E[(x_A R_A + x_B R_B) - E[x_A R_A + x_B R_B])^2] \\ &= E[(x_A(R_A - \mu_A) + x_B(R_B - \mu_B))^2] \\ &= E[x_A^2 (R_A - \mu_A)^2 + x_B^2 (R_B - \mu_B)^2 + 2 x_A x_B (R_A - \mu_A)(R_B - \mu_B)] \\ &= x_A^2 E[(R_A - \mu_A)^2] + x_B^2 E[(R_B - \mu_B)^2] + 2 x_A x_B E[(R_A - \mu_A)(R_B - \mu_B)], \end{aligned}$$

and the result follows by the de&nitions of $var(R_A), var(R_B)$ and $cov(R_A, R_B)$..

Notice that the variance of the portfolio is a weighted average of the variances of the individual assets plus two times the product of the portfolio weights times the covariance between the assets. If the portfolio weights are both positive then a positive covariance will tend to increase the portfolio variance, because both returns tend to move in the same direction, and a negative covariance will tend to reduce the portfolio variance. Thus &nding negatively correlated returns can be very bene&cial when forming portfolios. What is surprising is that a positive covariance can also be bene&cial to diversi&cation.

1.2 Efficient portfolios with two risky assets

In this section we describe how mean-variance efficient portfolios are constructed. First we make some assumptions:

Assumptions

- Returns are jointly normally distributed. This implies that means, variances and covariances of returns completely characterize the joint distribution of returns.
- Investors only care about portfolio expected return and portfolio variance. Investors like portfolios with high expected return but dislike portfolios with high return variance.

Given the above assumptions we set out to characterize the set of portfolios that have the highest expected return for a given level of risk as measured by portfolio variance. These portfolios are called efficient portfolios and are the portfolios that investors are most interested in holding.

For illustrative purposes we will show calculations using the data in the table below.

Table 1: Example Data

μ_A	μ_B	σ_A^2	σ_B^2	σ_A	σ_B	σ_{AB}	ρ_{AB}
0.175	0.055	0.067	0.013	0.258	0.115	-0.004875	-0.164

The collection of all feasible portfolios (the investment possibilities set) in the case of two assets is simply all possible portfolios that can be formed by varying the portfolio weights x_A and x_B such that the weights sum to one ($x_A + x_B = 1$). We summarize the expected return-risk (mean-variance) properties of the feasible portfolios in a plot with portfolio expected return, μ_p, on the vertical axis and portfolio *standard-deviation,* σ_p, on the horizontal axis. The portfolio standard deviation is used instead of variance because standard deviation is measured in the same units as the expected value (recall, variance is the average squared deviation from the mean).

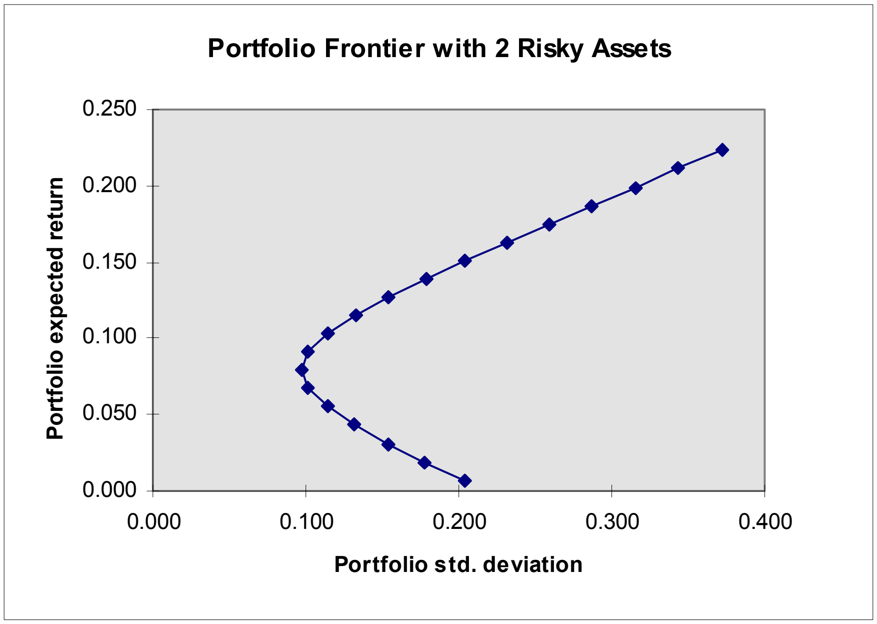


Figure 1

The investment possibilities set or portfolio frontier for the data in Table 1 is illustrated in Figure 1. Here the portfolio weight on asset A, x_A, is varied from -0.4 to 1.4 in increments of 0.1 and, since $x_B = 1 - x_A$, the weight on asset is then varies from 1.4 to -0.4. This gives us 18 portfolios with weights $(x_A, x_B) = (-0.4, 1.4), (-0.3, 1.3), ..., (1.3, -0.3), (1.4, -0.4)$. For each of these portfolios we use the formulas (2) and (3) to compute μ_p and $\sigma_p = \sqrt{\sigma_p^2}$. We then plot these values[1].

Notice that the plot in (μ_p, σ_p) space looks like a parabola turned on its side (in fact it is one side of a hyperbola). Since investors desire portfolios with the highest expected return for a given level of risk, combinations that are in the upper left corner are the best portfolios and those in the lower right corner are the worst. Notice that the portfolio at the bottom of the parabola has the property that it has the smallest variance among all feasible portfolios. Accordingly, this portfolio is called the *global minimum variance* portfolio.

It is a simple exercise in calculus to &nd the global minimum variance portfolio. We solve the constrained optimization problem

$$\begin{aligned} \min_{x_A, x_B} \ \sigma_p^2 &= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + 2x_Ax_B\sigma_{AB} \\ s.t. \ x_A + x_B &= 1. \end{aligned}$$

[1]The careful reader may notice that some of the portfolio weights are negative. A negative portfolio weight indicates that the asset is sold short and the proceeds of the short sale are used to buy more of the other asset. A short sale occurs when an investor borrows an asset and sells it in the market. The short sale is closed out when the investor buys back the asset and then returns the borrowed asset. If the asset price drops then the short sale produces and pro&t.

Substituting $x_B = 1 - x_A$ into the formula for σ_p^2 reduces the problem to

$$\min_{x_A} \ \sigma_p^2 = x_A^2\sigma_A^2 + (1 - x_A)^2\sigma_B^2 + 2x_A(1 - x_A)\sigma_{AB}.$$

The first order conditions for a minimum, via the chain rule, are

$$0 = \frac{d\sigma_p^2}{dx_A} = 2x_A^{\min}\sigma_A^2 - 2(1 - x_A^{\min})\sigma_B^2 + 2\sigma_{AB}(1 - 2x_A^{\min})$$

and straightforward calculations yield

$$x_A^{\min} = \frac{\sigma_B^2 - \sigma_{AB}}{\sigma_A^2 + \sigma_B^2 - 2\sigma_{AB}}, \ x_B^{\min} = 1 - x_A^{\min}. \tag{4}$$

For our example, using the data in table 1, we get $x_A^{\min} = 0.2$ and $x_B^{\min} = 0.8$.

Efficient portfolios are those with the highest expected return for a given level of risk. Inefficient portfolios are then portfolios such that there is another feasible portfolio that has the same risk (σ_p) but a higher expected return (μ_p). From the plot it is clear that the inefficient portfolios are the feasible portfolios that lie below the global minimum variance portfolio and the efficient portfolios are those that lie above the global minimum variance portfolio.

The shape of the investment possibilities set is very sensitive to the correlation between assets A and B. If ρ_{AB} is close to 1 then the investment set approaches a straight line connecting the portfolio with all wealth invested in asset B, $(x_A, x_B) = (0, 1)$, to the portfolio with all wealth invested in asset A, $(x_A, x_B) = (1, 0)$. This case is illustrated in Figure 2. As ρ_{AB} approaches zero the set starts to bow toward the μ_p axis and the power of diversification starts to kick in. If $\rho_{AB} = -1$ then the set actually touches the μ_p axis. What this means is that if assets A and B are perfectly negatively correlated then there exists a portfolio of A and B that has positive expected return and zero variance! To find the portfolio with $\sigma_p^2 = 0$ when $\rho_{AB} = -1$ we use (4) and the fact that $\sigma_{AB} = \rho_{AB}\sigma_A\sigma_B$ to give

$$x_A^{\min} = \frac{\sigma_B}{\sigma_A + \sigma_B}, \ x_B^{\min} = 1 - x_A$$

The case with $\rho_{AB} = -1$ is also illustrated in Figure 2.

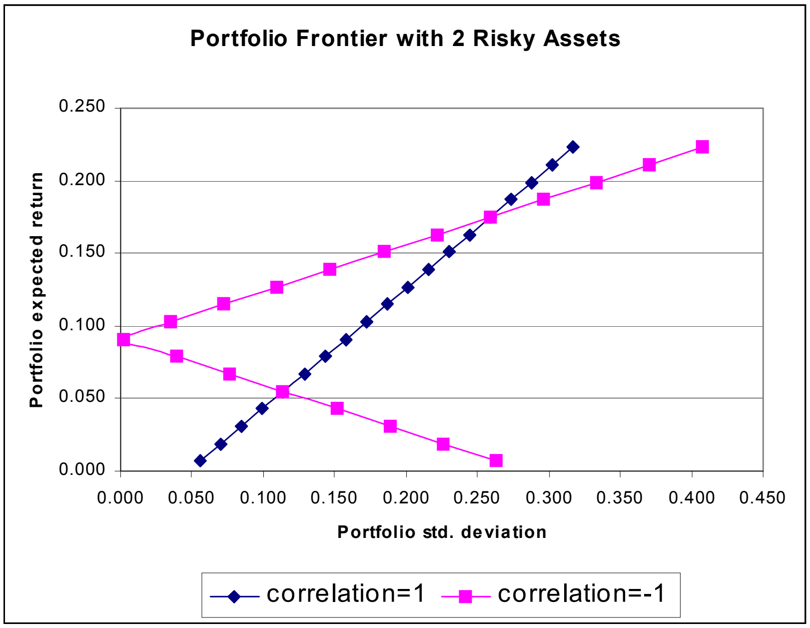


Figure 2

Given the efficient set of portfolios, which portfolio will an investor choose? Of the efficient portfolios, investors will choose the one that accords with their risk preferences. Very risk averse investors will choose a portfolio very close to the global minimum variance portfolio and very risk tolerant investors will choose portfolios with large amounts of asset A which may involve short-selling asset B.

1.3 Efficient portfolios with a risk-free asset

In the preceding section we constructed the efficient set of portfolios in the absence of a risk-free asset. Now we consider what happens when we introduce a risk free asset. In the present context, a risk free asset is equivalent to default-free pure discount bond that matures at the end of the assumed investment horizon. The risk-free rate, r_f, is then the return on the bond, assuming no in! ation. For example, if the investment horizon is one month then the risk-free asset is a 30-day Treasury bill (T-bill) and the risk free rate is the nominal rate of return on the T-bill. If our holdings of the risk free asset is positive then we are lending money at the risk-free rate and if our holdings are negative then we are borrowing at the risk-free rate.

1.3.1 Efficient portfolios with one risky asset and one risk free asset

Continuing with our example, consider an investment in asset B and the risk free asset (henceforth referred to as a T-bill) and suppose that $r_f = 0.03$. Since the risk free rate is &xed over the investment horizon it has some special properties, namely

$$\mu_f = E[r_f] = r_f$$

$$\begin{aligned} var(r_f) &= 0 \\ cov(R_B, r_f) &= 0 \end{aligned}$$

Let x_B denote the share of wealth in asset B and $x_f = 1 - x_B$ denote the share of wealth in T-bills. The portfolio expected return is

$$\begin{aligned} R_p &= x_B R_B + (1 - x_B) r_f \\ &= x_B (R_B - r_f) + r_f \end{aligned}$$

The quantity $R_B - r_f$ is called the *excess return* (over the return on T-bills) on asset B. The portfolio expected return is then

$$\mu_p = x_B(\mu_B - r_f) + r_t$$

where the quantity $(\mu_B - r_f)$ is called the *expected excess return* or *risk premium* on asset B. We may express the risk premium on the portfolio in terms of the risk premium on asset B:

$$\mu_p - r_f = x_B(\mu_B - r_f)$$

The more we invest in asset B the higher the risk premium on the portfolio.

The portfolio variance only depends on the variability of asset B and is given by

$$\sigma_p^2 = x_B^2 \sigma_B^2.$$

The portfolio standard deviation is therefore proportional to the standard deviation on asset B:

$$\sigma_p = x_B \sigma_B$$

which can use to solve for x_B

$$x_B = \frac{\sigma_p}{\sigma_B}$$

Using the last result, the feasible (and efficient) set of portfolios follows the equation

$$\mu_p = r_f + \frac{\mu_B - r_f}{\sigma_B} \cdot \sigma_p \tag{5}$$

which is simply straight line in (μ_p, σ_p) with intercept r_f and slope $\frac{\mu_B - r_f}{\sigma_B}$. The slope of the combination line between T-bills and a risky asset is called the *Sharpe ratio* or *Sharpe's slope* and it measures the risk premium on the asset per unit of risk (as measured by the standard deviation of the asset).

The portfolios which are combinations of asset A and T-bills and combinations of asset B and T-bills using the data in Table 1 with $r_f = 0.03$. is illustrated in Figure 4.

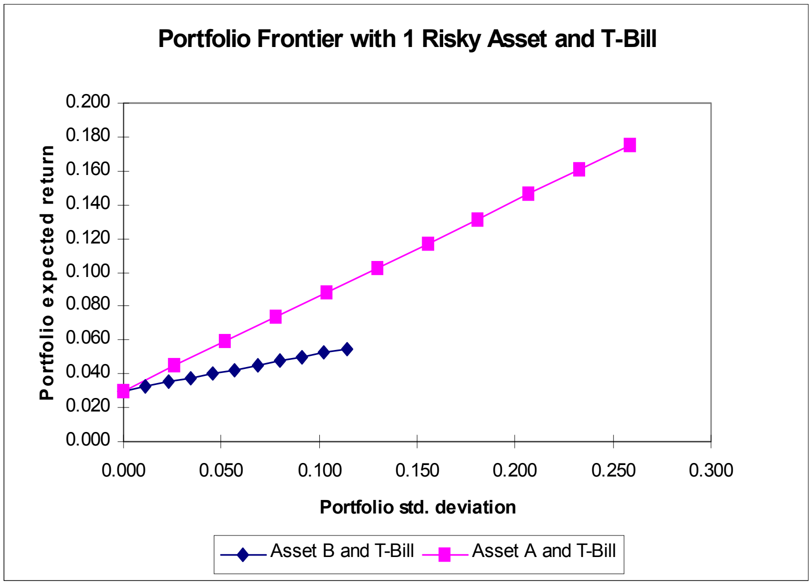


Figure 3

Notice that expected return-risk trade off of these portfolios is linear. Also, notice that the portfolios which are combinations of asset A and T-bills have expected returns uniformly higher than the portfolios consisting of asset B and T-bills. This occurs because the Sharpe's slope for asset A is higher than the slope for asset B:

$$\frac{\mu_A - r_f}{\sigma_A} = \frac{0.175 - 0.03}{0.258} = 0.562, \quad \frac{\mu_B - r_f}{\sigma_B} = \frac{0.055 - 0.03}{0.115} = 0.217.$$

Hence, portfolios of asset A and T-bills are efficient relative to portfolios of asset B and T-bills.

1.3.2 Efficient portfolios with two risky assets and a risk-free asset

Now we expand on the previous results by allowing our investor to form portfolios of assets A, B and T-bills. The efficient set in this case will still be a straight line in $(\mu_p, \sigma_p)-$ space with intercept r_f. The slope of the efficient set, the maximum Sharpe ratio, is such that it is tangent to the efficient set constructed just using the two risky assets A and B. Figure 5 illustrates why this is so.

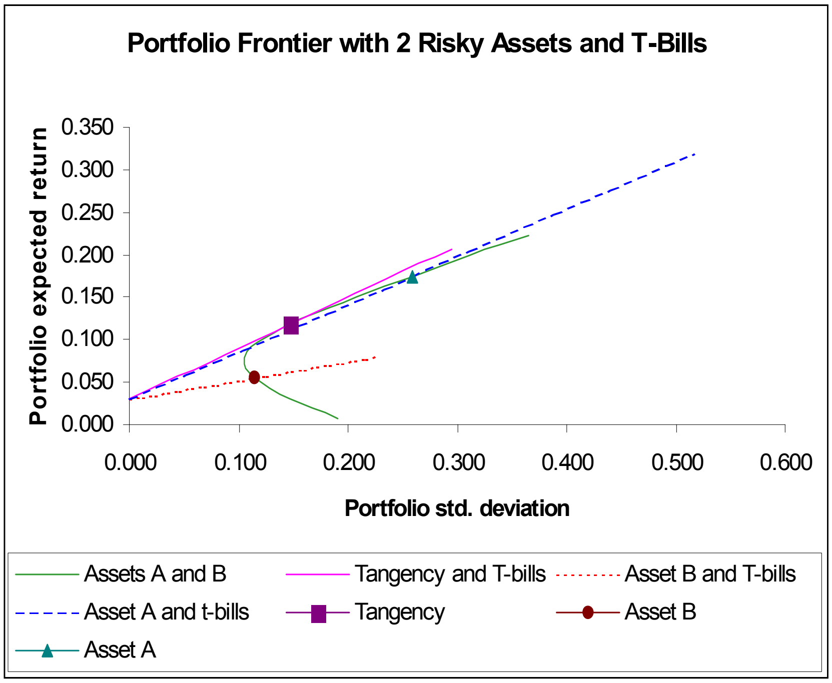


Figure 4

If we invest in only in asset B and T-bills then the Sharpe ratio is $\frac{\mu_B - r_f}{\sigma_B} = 0.217$ and the CAL intersects the parabola at point B. This is clearly not the efficient set of portfolios. For example, we could do uniformly better if we instead invest only in asset A and T-bills. This gives us a Sharpe ratio of $\frac{\mu_A - r_f}{\sigma_A} = 0.562$ and the new CAL intersects the parabola at point A. However, we could do better still if we invest in T-bills and some combination of assets A and B. Geometrically, it is easy to see that the best we can do is obtained for the combination of assets A and B such that the CAL is just tangent to the parabola. This point is marked T on the graph and represents the *tangency portfolio* of assets A and B.

We can determine the proportions of each asset in the tangency portfolio by &nding the values of x_A and x_B that maximize the Sharpe ratio of a portfolio that is on the envelope of the parabola. Formally, we solve

$$\max_{x_A, x_B} \frac{\mu_p - r_f}{\sigma_p} \quad s.t.$$

$$\begin{aligned} \mu_p &= x_A \mu_A + x_B \mu_B \\ \sigma_p^2 &= x_A^2 \sigma_A^2 + x_B^2 \sigma_B^2 + 2 x_A x_B \sigma_{AB} \\ 1 &= x_A + x_B \end{aligned}$$

After various substitutions, the above problem can be reduced to

$$\max_{x_A} \frac{x_A(\mu_A - r_f) + (1 - x_A)(\mu_B - r_f)}{(x_A^2 \sigma_A^2 + (1 - x_A)^2 \sigma_B^2 + 2 x_A (1 - x_A) \sigma_{AB})^{1/2}}.$$

This is a straightforward, albeit very tedious, calculus problem and the solution can be shown to be

$$x_A^T = \frac{(\mu_A - r_f)\sigma_B^2 - (\mu_B - r_f)\sigma_{AB}}{(\mu_A - r_f)\sigma_B^2 + (\mu_B - r_f)\sigma_A^2 - (\mu_A - r_f + \mu_B - r_f)\sigma_{AB}},\ x_B^T = 1 - x_A^T.$$

For the example data using $r_f = 0.03$, we get $x_A^T = 0.542$ and $x_B^T = 0.458$. The expected return on the tangency portfolio is

$$\begin{aligned}\mu_T &= x_A^T\mu_A + x_B^T\mu_B \\ &= (0.542)(0.175) + (0.458)(0.055) = 0.110,\end{aligned}$$

the variance of the tangency portfolio is

$$\begin{aligned}\sigma_T^2 &= \left(x_A^T\right)^2\sigma_A^2 + \left(x_B^T\right)^2\sigma_B^2 + 2x_A^Tx_B^T\sigma_{AB} \\ &= (0.542)^2(0.067) + (0.458)^2(0.013) + 2(0.542)(0.458) = 0.015,\end{aligned}$$

and the standard deviation of the tangency portfolio is

$$\sigma_T = \sqrt{\sigma_T^2} = \sqrt{0.015} = 0.124.$$

The efficient portfolios now are combinations of the tangency portfolio and the T-bill. This important result is known as the *mutual fund separation theorem*. The tangency portfolio can be considered as a mutual fund of the two risky assets, where the shares of the two assets in the mutual fund are determined by the tangency portfolio weights, and the T-bill can be considered as a mutual fund of risk free assets. The expected return-risk trade-off of these portfolios is given by the line connecting the risk-free rate to the tangency point on the efficient frontier of risky asset only portfolios. Which combination of the tangency portfolio and the T-bill an investor will choose depends on the investor's risk preferences. If the investor is very risk averse, then she will choose a combination with very little weight in the tangency portfolio and a lot of weight in the T-bill. This will produce a portfolio with an expected return close to the risk free rate and a variance that is close to zero.

For example, a highly risk averse investor may choose to put 10% of her wealth in the tangency portfolio and 90% in the T-bill. Then she will hold $(10\%) \times (54.2\%) = 5.42\%$ of her wealth in asset A, $(10\%) \times (45.8\%) = 4.58\%$ of her wealth in asset B and 90% of her wealth in the T-bill. The expected return on this portfolio is

$$\begin{aligned}\mu_p &= r_f + 0.10(\mu_T - r_f) \\ &= 0.03 + 0.10(0.110 - 0.03) \\ &= 0.038.\end{aligned}$$

and the standard deviation is

$$\begin{aligned}\sigma_p &= 0.10\sigma_T \\ &= 0.10(0.124) \\ &= 0.012.\end{aligned}$$

A very risk tolerant investor may actually borrow at the risk free rate and use these funds to leverage her investment in the tangency portfolio. For example, suppose the risk tolerant investor borrows 10% of her wealth at the risk free rate and uses the proceed to purchase 110% of her wealth in the tangency portfolio. Then she would hold $(110\%)\times(54.2\%) = 59.62\%$ of her wealth in asset A, $(110\%)\times(45.8\%) = 50.38\%$ in asset B and she would owe 10% of her wealth to her lender. The expected return and standard deviation on this portfolio is

$$\begin{aligned}\mu_p &= 0.03 + 1.1(0.110 - 0.03) = 0.118 \\ \sigma_p &= 1.1(0.124) = 0.136.\end{aligned}$$

2 Efficient Portfolios and Value-at-Risk

As we have seen, efficient portfolios are those portfolios that have the highest expected return for a given level of risk as measured by portfolio standard deviation. For portfolios with expected returns above the T-bill rate, efficient portfolios can also be characterized as those portfolios that have minimum risk (as measured by portfolio standard deviation) for a given target expected return.

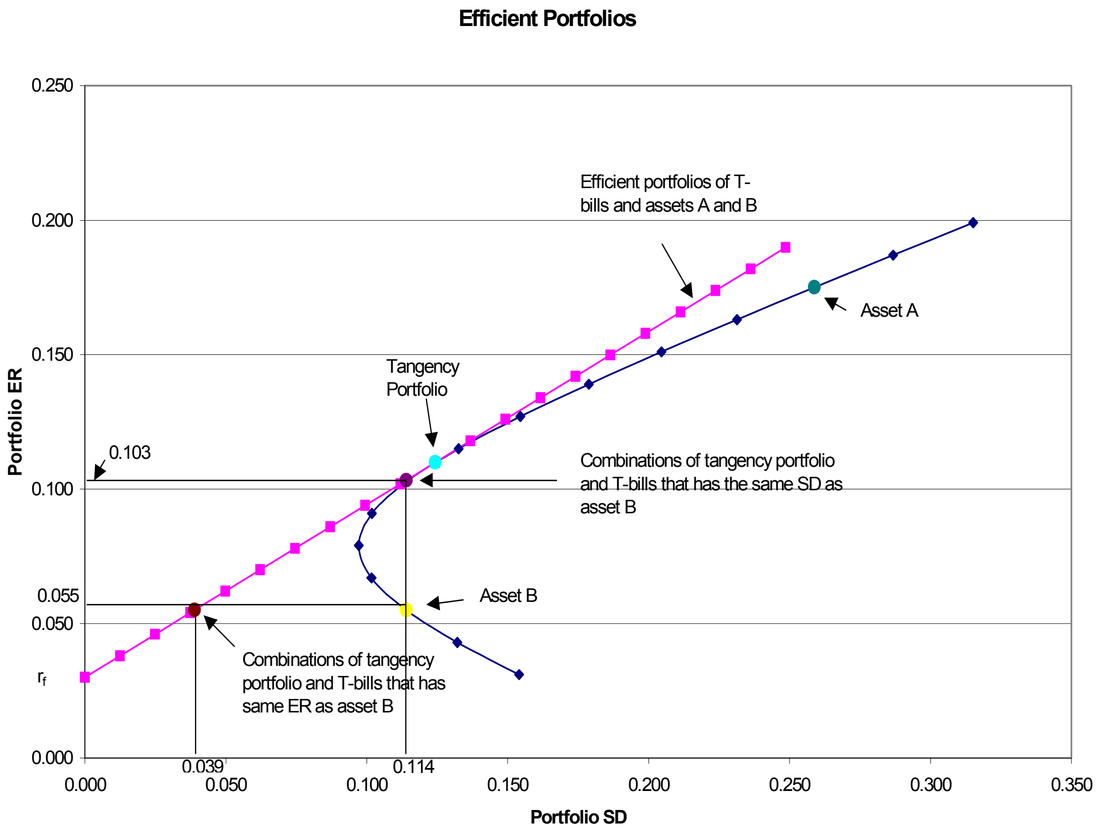


Figure 5

To illustrate, consider &gure 5 which shows the portfolio frontier for two risky assets and the efficient frontier for two risky assets plus a risk-free asset. Suppose an investor initially holds all of his wealth in asset A. The expected return on this portfolio is $\mu_B = 0.055$ and the standard deviation (risk) is $\sigma_B = 0.115$. An efficient portfolio (combinations of the tangency portfolio and T-bills) that has the same standard deviation (risk) as asset B is given by the portfolio on the efficient frontier that is directly above $\sigma_B = 0.115$. To &nd the shares in the tangency portfolio and T-bills in this portfolio recall from (xx) that the standard deviation of a portfolio with x_T invested in the tangency portfolio and $1 - x_T$ invested in T-bills is $\sigma_p = x_T \sigma_T$. Since we want to &nd the efficient portfolio with $\sigma_p = \sigma_B = 0.115$, we solve

$$x_T = \frac{\sigma_B}{\sigma_T} = \frac{0.115}{0.124} = 0.917, \ x_f = 1 - x_T = 0.083.$$

That is, if we invest 91.7% of our wealth in the tangency portfolio and 8.3% in T-bills we will have a portfolio with the same standard deviation as asset B. Since this is an efficient portfolio, the expected return should be higher than the expected return on asset B. Indeed it is since

$$\begin{aligned}\mu_p &= r_f + x_T(\mu_T - r_f) \\ &= 0.03 + 0.917(0.110 - 0.03) \\ &= 0.103\end{aligned}$$

Notice that by diversifying our holding into assets A, B and T-bills we can obtain a portfolio with the same risk as asset B but with almost twice the expected return!

Next, consider &nding an efficient portfolio that has the same expected return as asset B. Visually, this involves &nding the combination of the tangency portfolio and T-bills that corresponds with the intersection of a horizontal line with intercept $\mu_B = 0.055$ and the line representing efficient combinations of T-bills and the tangency portfolio. To &nd the shares in the tangency portfolio and T-bills in this portfolio recall from (xx) that the expected return of a portfolio with x_T invested in the tangency portfolio and $1 - x_T$ invested in T-bills has expected return equal to $\mu_p = r_f + x_T(\mu_T - r_f)$. Since we want to &nd the efficient portfolio with $\mu_p = \mu_B = 0.055$ we use the relation

$$\mu_p - r_f = x_T(\mu_T - r_F)$$

and solve for x_T and $x_f = 1 - x_T$

$$x_T = \frac{\mu_p - r_f}{\mu_T - r_f} = \frac{0.055 - 0.03}{0.110 - 0.03} = 0.313, x_f = 1 - x_T = 0.687.$$

That is, if we invest 31.3% of wealth in the tangency portfolio and 68.7% of our wealth in T-bills we have a portfolio with the same expected return as asset B. Since this is an efficient portfolio, the standard deviation (risk) of this portfolio should be lower than the standard deviation on asset B. Indeed it is since

$$\begin{aligned}\sigma_p &= x_T\sigma_T \\ &= 0.313(0.124) \\ &= 0.039.\end{aligned}$$

Notice how large the risk reduction is by forming an efficient portfolio. The standard deviation on the efficient portfolio is almost three times smaller than the standard deviation of asset B!

The above example illustrates two ways to interpret the bene&ts from forming efficient portfolios. Starting from some benchmark portfolio, we can &x standard deviation (risk) at the value for the benchmark and then determine the gain in expected return from forming a diversi&ed portfolio[2]. The gain in expected return has concrete

[2] The gain in expected return by investing in an efficient portfolio abstracts from the costs associated with selling the benchmark portfolio and buying the efficient portfolio.

meaning. Alternatively, we can &x expected return at the value for the benchmark and then determine the reduction in standard deviation (risk) from forming a diversi&ed portfolio. The meaning to an investor of the reduction in standard deviation is not as clear as the meaning to an investor of the increase in expected return. It would be helpful if the risk reduction bene&t can be translated into a number that is more interpretable than the standard deviation. The concept of Value-at-Risk (VaR) provides such a translation.

Recall, the VaR of an investment is the expected loss in investment value over a given horizon with a stated probability. For example, consider an investor who invests $W_0 = \$100,000$ in asset B over the next year. Assume that R_B represents the annual (continuously compounded) return on asset B and that $R_B\ \tilde{}\ N(0.055, (0.114)^2)$. The 5% annual VaR of this investment is the loss that would occur if return on asset B is equal to the 5% left tail quantile of the normal distribution of R_B. The 5% quantile, $q_{0.05}$ is determined by solving

$$\Pr(R_B \leq q_{0.05}) = 0.05.$$

Using the inverse cdf for a normal random variable with mean 0.055 and standard deviation 0.114 it can be shown that $q_{0.05} = -0.133$.That is, with 5% probability the return on asset B will be -13.3% or less. If $R_B = -0.133$ then the loss in portfolio value[3], which is the 5% VaR, is

$$\text{loss in portfolio value} = VaR = |W_0 \cdot (e^{q_{0.05}} - 1)| = |\$100,000(e^{-0.133} - 1)| = \$12,413.$$

To reiterate, if the investor hold \$100,000 in asset B over the next year then the 5% VaR on the portfolio is $\$12,413$. This is the loss that would occur with 5% probability.

Now suppose the investor chooses to hold an efficient portfolio with the same expected return as asset B. This portfolio consists of 31.3% in the tangency portfolio and 68.7% in T-bills and has a standard deviation equal to 0.039. Let R_p denote the annual return on this portfolio and assume that $R_p\ \tilde{}\ N(0.055, 0.039)$. Using the inverse cdf for this normal distribution, the 5% quantile can be shown to be $q_{0.05} = -0.009$. That is, with 5% probability the return on the efficient portfolio will be -0.9% or less. This is considerably smaller than the 5% quantile of the distribution of asset B. If $R_p = -0.009$ the loss in portfolio value (5% VaR) is

$$\text{loss in portfolio value} = VaR = |W_0 \cdot (e^{q_{0.05}} - 1)| = |\$100,000(e^{-0.009} - 1)| = \$892.$$

Notice that the 5% VaR for the efficient portfolio is almost &fteen times smaller than the 5% VaR of the investment in asset B. Since VaR translates risk into a dollar &gure it is more interpretable than standard deviation.

[3] To compute the VaR we need to convert the continuous compounded return (quantile) to a simple return (quantile). Recall, if R_t^c is a continuously compounded return and R_t is a somple return then $R_t^c = \ln(1 + R_t)$ and $R_t = e^{R_t^c} - 1$.

3 Further Reading

The classic text on portfolio optimization is Markowitz (1954). Good intermediate level treatments are given in Benninga (2000), Bodie, Kane and Marcus (1999) and Elton and Gruber (1995). An interesting recent treatment with an emphasis on statistical properties is Michaud (1998). Many practical results can be found in the *Financial Analysts Journal* and the *Journal of Portfolio Management*. An excellent overview of value at risk is given in Jorian (1997).

4 Appendix Review of Optimization and Constrained Optimization

Consider the function of a single variable

$$y = f(x) = x^2$$

which is illustrated in Figure xxx. Clearly the minimum of this function occurs at the point $x = 0$. Using calculus, we &nd the minimum by solving

$$\min_x \; y = x^2.$$

The &rst order (necessary) condition for a minimum is

$$0 = \frac{d}{dx} f(x) = \frac{d}{dx} x^2 = 2x$$

and solving for x gives $x = 0$. The second order condition for a minimum is

$$0 < \frac{d^2}{dx} f(x)$$

and this condition is clearly satis&ed for $f(x) = x^2$.

Next, consider the function of two variables

$$y = f(x, z) = x^2 + z^2 \tag{6}$$

which is illustrated in Figure xxx.

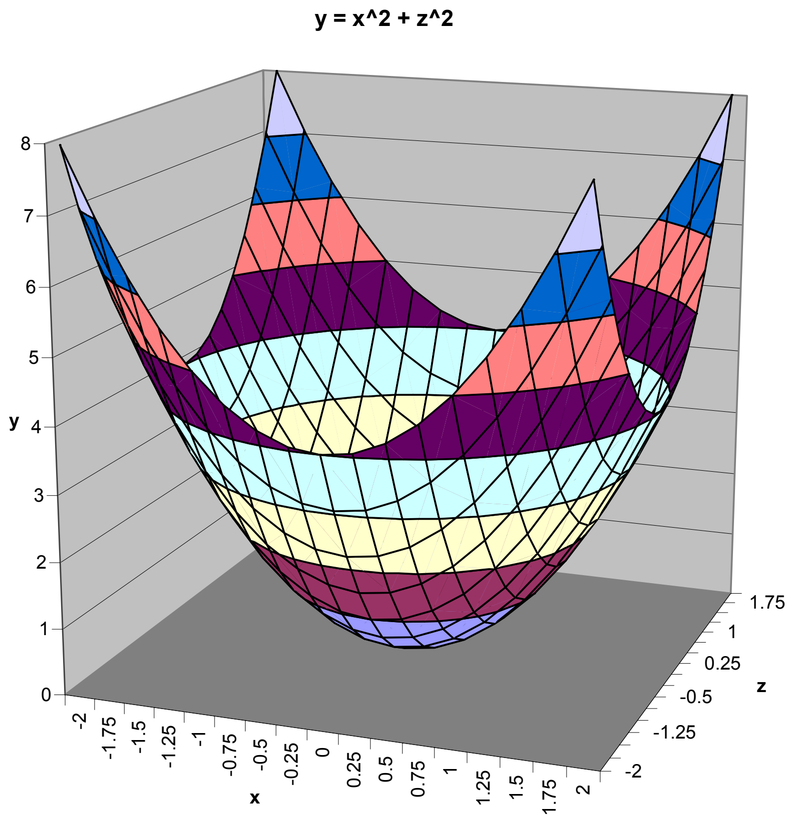


Figure 6

This function looks like a salad bowl whose bottom is at $x = 0$ and $z = 0$. To find the minimum of (6), we solve

$$\min_{x,z} y = x^2 + z^2$$

and the first order necessary conditions are

$$0 = \frac{\partial y}{\partial x} = 2x$$

and

$$0 = \frac{\partial y}{\partial z} = 2z.$$

Solving these two equations gives $x = 0$ and $z = 0$.

Now suppose we want to minimize (6) subject to the linear constraint

$$x + z = 1. \tag{7}$$

The minimization problem is now a *constrained minimization*

$$\begin{aligned} \min_{x,z} y &= x^2 + z^2 \text{ subject to } (s.t.) \\ x + z &= 1 \end{aligned}$$

and is illustrated in Figure xxx. Given the constraint $x + z = 1$, the function (6) is no longer minimized at the point $(x, z) = (0, 0)$ because this point does not satisfy $x + z = 1$. The One simple way to solve this problem is to substitute the restriction (7) into the function (6) and reduce the problem to a minimization over one variable. To illustrate, use the restriction (7) to solve for z as

$$z = 1 - x. \tag{8}$$

Now substitute (7) into (6) giving

$$y = f(x, z) = f(x, 1 - x) = x^2 + (1 - x)^2. \tag{9}$$

The function (9) satis&es the restriction (7) by construction. The constrained minimization problem now becomes

$$\min_x y = x^2 + (1 - x)^2.$$

The &rst order conditions for a minimum are

$$0 = \frac{d}{dx}(x^2 + (1 - x)^2) = 2x - 2(1 - x) = 4x - 2$$

and solving for x gives $x = 1/2$. To solve for z, use (8) to give $z = 1 - (1/2) = 1/2$. Hence, the solution to the constrained minimization problem is $(x, z) = (1/2, 1/2)$.

Another way to solve the constrained minimization is to use the method of *Lagrange multipliers*. This method augments the function to be minimized with a linear function of the constraint in homogeneous form. The constraint (7) in homogenous form is

$$x + z - 1 = 0$$

The augmented function to be minimized is called the *Lagrangian* and is given by

$$L(x, z, \lambda) = x^2 + z^2 - \lambda(x + z - 1).$$

The coefficient on the constraint in homogeneous form, λ, is called the Lagrange multiplier. It measures the cost, or shadow price, of imposing the constraint relative to the unconstrained problem. The constrained minimization problem to be solved is now

$$\min_{x,z,\lambda} \; L(x, z, \lambda) = x^2 + z^2 + \lambda(x + z - 1).$$

The &rst order conditions for a minimum are

$$\begin{aligned}
0 &= \frac{\partial L(x, z, \lambda)}{\partial x} = 2x + \lambda \\
0 &= \frac{\partial L(x, z, \lambda)}{\partial z} = 2z + \lambda \\
0 &= \frac{\partial L(x, z, \lambda)}{\partial \lambda} = x + z - 1
\end{aligned}$$

The &rst order conditions give three linear equations in three unknowns. Notice that the &rst order condition with respect to λ imposes the constraint. The &rst two conditions give

$$2x = 2z = -\lambda$$

or

$$x = z.$$

Substituting $x = z$ into the third condition gives

$$2z - 1 = 0$$

or

$$z = 1/2.$$

The &nal solution is $(x, y, \lambda) = (1/2, 1/2, -1)$.

The Lagrange multiplier, λ, measures the marginal cost, in terms of the value of the objective function, of imposing the constraint. Here, $\lambda = -1$ which indicates that imposing the constraint $x + z = 1$ reduces the objective function. To understand the roll of the Lagrange multiplier better, consider imposing the constraint $x + z = 0$. Notice that the unconstrained minimum achieved at $x = 0, z = 0$ satis&es this constraint. Hence, imposing $x + z = 0$ does not cost anything and so the Lagrange multiplier associated with this constraint should be zero. To con&rm this, the we solve the problem

$$\min_{x,z,\lambda} \; L(x, z, \lambda) = x^2 + z^2 + \lambda(x + z - 0).$$

The &rst order conditions for a minimum are

$$\begin{aligned} 0 &= \frac{\partial L(x, z, \lambda)}{\partial x} = 2x - \lambda \\ 0 &= \frac{\partial L(x, z, \lambda)}{\partial z} = 2z - \lambda \\ 0 &= \frac{\partial L(x, z, \lambda)}{\partial \lambda} = x + z \end{aligned}$$

The &rst two conditions give

$$2x = 2z = -\lambda$$

or

$$x = z.$$

Substituting $x = z$ into the third condition gives

$$2z = 0$$

or

$$z = 0.$$

The &nal solution is $(x, y, \lambda) = (0, 0, 0)$. Notice that the Lagrange multiplier, λ, is equal to zero in this case.

5 Problems

Exercise 1 *Consider the problem of investing in two risky assets A and B and a risk-free asset (T-bill). The optimization problem to find the tangency portfolio may be reduced to*

$$\max_{x_A} \frac{x_A(\mu_A - r_f) + (1 - x_A)(\mu_B - r_f)}{\left(x_A^2\sigma_A^2 + (1 - x_A)^2\sigma_B^2 + 2x_A(1 - x_A)\sigma_{AB}\right)^{1/2}}$$

where x_A is the share of wealth in asset A in the tangency portfolio and $x_B = 1 - x_A$ is the share of wealth in asset B in the tangency portfolio. Using simple calculus, show that

$$x_A = \frac{(\mu_A - r_f)\sigma_B^2 - (\mu_B - r_f)\sigma_{AB}}{(\mu_A - r_f)\sigma_B^2 + (\mu_B - r_f)\sigma_A^2 - (\mu_A - r_f + \mu_B - r_f)\sigma_{AB}}.$$

References

[1] Benninga, S. (2000), *Financial Modeling*, Second Edition. Cambridge, MA: MIT Press.

[2] Bodie, Kane and Marcus (199x), *Investments*, xxx Edition.

[3] Elton, E. and G. Gruber (1995). *Modern Portfolio Theory and Investment Analysis*, Fifth Edition. New York: Wiley.

[4] Jorian, P. (1997). *Value at Risk*. New York: McGraw-Hill.

[5] Markowitz, H. (1987). *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Cambridge, MA: Basil Blackwell.

[6] Markowitz, H. (1991). *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley, 1959; 2nd ed., Cambridge, MA: Basil Blackwell.

[7] Michaud, R.O. (1998). *Efficient Asset Management: A Practical Guide to Stock Portfolio Optimization and Asset Allocation*. Boston, MA:Harvard Business School Press.

Introduction to Financial Econometrics Chapter 5 The Markowitz Algorithm

Eric Zivot
Department of Economics
University of Washington

January 26, 2000
This version: February 19, 2000

1 Efficient Portfolios with Three Risky Assets: The Markowitz Algorithm

Consider the portfolio problem with three risky assets denoted A, B and C. Let R_i $(i = A, B, C)$ denote the return on asset i and assume that

$$\begin{aligned} R_i &\sim i.i.d.\ N(\mu_i, \sigma_i^2) \\ cov(R_i, R_j) &= \sigma_{ij}. \end{aligned}$$

For illustrative purposes, Table 1 provides example data on means, variances and covariances.

Table 1

Stock	μ_i	σ_i^2	Pari (i,j)	σ_{ij}
A	0.229	0.924	(A,B)	0.063
B	0.138	0.862	(A,C)	-0.582
C	0.052	0.528	(B,C)	-0.359

Let x_i denote the share of wealth invested in asset i and assume that all wealth is invested in the three assets so that $x_A + x_B + x_C = 1$. The portfolio return, R_p, is the random variable

$$R_{p,x} = x_A R_A + x_B R_B + x_C R_C.$$

The subscript "x" indicates that the portfolio is constructed using the x-weights x_A, x_B and x_C. The expected return on the portfolio is

$$\mu_{p,x} = E[R_{p,x}] = x_A\mu_A + x_B\mu_B + x_C\mu_C \tag{1}$$

and the variance of the portfolio return is

$$\sigma^2_{p,x} = var(R_{p,x}) = x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC}. \quad (2)$$

Notice that variance of the portfolio return depends on three variance terms and six covariance terms. Hence, with three assets there are twice as many covariance terms than variance terms contributing to portfolio variance. For example, let $x_i = 1/3$. Then

$$\begin{aligned}
\mu_{p,x} &= (\frac{1}{3})(0.229) + (\frac{1}{3})(0.138) + (\frac{1}{3})(0.528) \\
&= 0.140 \\
\sigma^2_{p,x} &= (\frac{1}{3})^2(0.924) + (\frac{1}{3})(0.862) + (\frac{1}{3})(0.528) \\
&\quad +2(\frac{1}{3})(\frac{1}{3})(0.063) + 2(\frac{1}{3})(\frac{1}{3})(-0.582) + 2(\frac{1}{3})(\frac{1}{3})(-0.359) \\
&= 0.062
\end{aligned}$$

The investment opportunity set is the set of portfolio expected return and portfolio standard deviation values for all possible portfolios such that $x_A + x_B + x_C = 1$. As in the two risky asset case, this set can be described in a graph with μ_p on the vertical axis and σ_p on the horizontal axis. Unlike the two asset case, however, the investment opportunity set cannot be simply described by one side of an hyperbola. The general shape of the set is complicated and depends crucially on the covariance terms σ_{ij}. As we shall see, we do not have to fully characterize the investment opportunity set. If we assume that investors only care about maximizing portfolio expected return and minimizing portfolio variance in deciding their asset allocation then we can simplify the portfolio problem by only concentrating on the combination of efficient portfolios between assets A, B and C. This is the framework originally developed by Harry Markowitz, the father of portfolio theory and winner of the Nobel Prize in economics.

We assume that the investor wishes to find portfolios that have the best expected return-risk trade-off. In other words, we assume that the investor seeks to find portfolios that maximize portfolio expected return for a given level of risk as measured by portfolio variance. Let $\sigma^2_{p,0}$ denote a target level of risk. Then the investor seeks to solve the constrained maximization problem

$$\begin{aligned}
\max_{x_A,x_B,x_C} \mu_{p,x} &= x_A\mu_A + x_B\mu_B + x_C\mu_C \text{ subject to } (s.t.) \qquad (3)\\
\sigma^2_{p,0} &= \sigma^2_{p,x} \\
&= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC} \\
1 &= x_A + x_B + x_C
\end{aligned}$$

This problem is illustrated in Figure xxx. The portfolio with weights (x_A, x_B, x_C) that satisfies the above maximization problem is, by definition, an efficient portfolio. The efficient portfolio frontier is graph of μ_p versus σ_p for the set of efficient portfolios generated by solving (3) for all possible target risk levels $\sigma^2_{p,0}$.[1] Just as in the two asset case, the efficient frontier in resembles one side of an hyperbola.

The investor's problem of maximizing portfolio expected return subject to a target level of risk has an equivalent dual representation in which the investor minimizes the risk of the portfolio (as measured by portfolio variance) subject to a target expected return level. Let $\mu_{p,0}$ denote a target expected return level. Then the dual problem is the constrained minimization problem

$$\begin{aligned} \min_{x_A,x_B,x_C} \sigma^2_{p,x} &= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 \\ &\quad +2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC} \quad s.t. \\ \mu_{p,0} &= x_A\mu_A + x_B\mu_B + x_C\mu_C \\ 1 &= x_A + x_B + x_C \end{aligned} \tag{4}$$

To find efficient portfolios of risky assets in practice, the dual problem (4) is most often solved. This is partially due to computational conveniences and partly due to investors being more willing to specify target expected returns rather than target risk levels. To solve the constrained minimization problem (4), we form the Lagrangian

$$\begin{aligned} L(x_A, x_B, x_C, \lambda_1, \lambda_2) &= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC} \\ &\quad +\lambda_1(x_A\mu_A + x_B\mu_B + x_C\mu_C - \mu_{p,0}) + \lambda_2(x_A + x_B + x_C - 1). \end{aligned}$$

The first order conditions for a minimum are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial x_A} = 2x_A\sigma_A^2 + 2x_B\sigma_{AB} + 2x_C\sigma_{AB} + \lambda_1\mu_A + \lambda_2 \\ 0 &= \frac{\partial L}{\partial x_A} = 2x_B\sigma_B^2 + 2x_A\sigma_{AB} + 2x_C\sigma_{BC} + \lambda_1\mu_B + \lambda_2 \\ 0 &= \frac{\partial L}{\partial x_A} = 2x_C\sigma_C^2 + 2x_A\sigma_{AC} + 2x_B\sigma_{BC} + \lambda_1\mu_C + \lambda_2 \\ 0 &= \frac{\partial L}{\partial \lambda_1} = x_A\mu_A + x_B\mu_B + x_C\mu_C - \mu_{p,0} \\ 0 &= \frac{\partial L}{\partial \lambda_2} = x_A + x_B + x_C - 1 \end{aligned} \tag{5}$$

These are five linear equations in five unknowns and a unique solution can be found as long as there are no linear dependencies among the equations. The solution for x_A, x_B and x_C gives an efficient portfolio with expected return $\mu_{p,x} = \mu_{p,0}$, variance $\sigma^2_{p,x}$ given by (2) and standard deviation $\sigma_{p,x}$. The pair $(\mu_{p,x}, \sigma_{p,x})$ plots as a single point on the efficient frontier of portfolios of three risky assets.

For example, using the data in Table 1 and a target expected return of $\mu_{p,0} = 0.01$ the solution for the efficient portfolio can be shown to be $x_A = -0.398, x_B = 0.331$

[1] Not all target risk levels are feasible. The feasible risk levels are those that are greater than or equal to the global minimum variance portfolio.

and $x_C = 1.067$. For future reference, call this portfolio "asset X". Notice that asset A is sold short in this portfolio. The expected return, variance and standard deviation of this portfolio are

$$\begin{aligned}
\mu_{p,x} &= \mu_{p,0} = (-0.398)(0.229) + (0.331)(0.138) + (1.067)(0.528) \\
&= 0.05 \\
\sigma^2_{p,x} &= (-0.398)^2(0.924) + (0.331)(0.862) + (1.067)(0.528) \\
&\quad +2(-0.398)(0.331)(0.063) + 2(-0.398)(1.067)(-0.582) + 2(0.331)(1.067)(-0.359) \\
&= 1.066 \\
\sigma_{p,x} &= \sqrt{1.066} = 1.033.
\end{aligned}$$

The pair $(\mu_{p,0}, \sigma_{p,x}) = (0.01, 1.033)$ is illustrated in figure xxx.

To get another point on the efficient frontier the minimization problem (4) needs to be solved using another target expected return value $\mu_{p,1} \neq \mu_{p,0}$. That is, we need to find a portfolio with weights y_A, y_B and y_C that solves

$$\begin{aligned}
\min_{y_A, y_B, y_C} \ \sigma^2_{p,y} &= y_A^2\sigma_A^2 + y_B^2\sigma_B^2 + y_C^2\sigma_C^2 + 2y_Ay_B\sigma_{AB} \\
&\quad +2y_Ay_C\sigma_{AC} + 2y_By_C\sigma_{BC} \quad s.t. \\
\mu_{p,1} &= y_A\mu_A + y_B\mu_B + y_C\mu_C \\
1 &= y_A + y_B + y_C
\end{aligned} \tag{6}$$

The solution for y_A, y_B and y_C gives an efficient portfolio with expected return $\mu_{p,y} = \mu_{p,1}$, variance $\sigma^2_{p,y}$ given by (??) and standard deviation $\sigma_{p,y}$. The pair $(\mu_{p,y}, \sigma_{p,y})$ plots as a single point different from $(\mu_{p,x}, \sigma_{p,y})$ on the efficient frontier of portfolios.

For example, using the data in Table 1 and a target expected return of $\mu_{p,0} = 0.25$ the solution for the efficient portfolio can be shown to be $x_A = 1.097, x_B = 0.045$ and $x_C = -0.142$. For future reference, call this portfolio "asset Y". Notice that asset C is sold short in this portfolio. The expected return, variance and standard deviation of this portfolio are

$$\begin{aligned}
\mu_{p,y} &= \mu_{p,1} = (1.097)(0.229) + (0.045)(0.138) + (-0.142)(0.528) \\
&= 0.25 \\
\sigma^2_{p,y} &= (1.097)^2(0.924) + (0.045)(0.862) + (-0.142)(0.528) \\
&\quad +2(1.097)(0.045)(0.063) + 2(1.097)(-0.142)(-0.582) + 2(0.045)(-0.142)(-0.359) \\
&= 1.316 \\
\sigma_{p,y} &= \sqrt{1.316} = 1.147.
\end{aligned}$$

The pair $(\mu_{p,1}, \sigma_{p,y}) = (0.25, 1.147)$ is illustrated in figure xxx.

To create the entire efficient frontier we could solve the minimization problem (4) for all possible target expected returns within some range. This brute force approach, while illustrative, is not very practical computationally. Fortunately, there is an easier way to compute the entire efficient frontier that only requires solving (4) for two target returns. As we shall see, given any two portfolios on the efficient frontier another portfolio on the efficient frontier is a simple convex combination of these two portfolios. Hence, the results for the construction of efficient portfolios with two risky assets can be used to compute efficient portfolios with an arbitrary number of risky assets.

To illustrate this result, consider the two efficient portfolios that are the solutions of (4) and (6). Now consider forming a new portfolio that is a convex combination of these two portfolios. Let z_x denote the share of wealth invested in asset X (first efficient portfolio) and let z_y denote the share of wealth invested in asset Y (second efficient portfolio) and impose the constraint $z_x + z_y = 1$. The expected return and variance of this portfolio is

$$\begin{aligned} \mu_{p,z} &= z_x\mu_{p,x} + z_y\mu_{p,y} && (7) \\ \sigma^2_{p,z} &= z_x^2\sigma^2_{p,x} + z_y^2\sigma^2_{p,y} + 2z_xz_y\sigma_{xy} && (8) \end{aligned}$$

where

$$\sigma_{xy} = cov(R_{p,x}, R_{p,y})$$

and $R_{p,x}$ denotes the return on asset X and $R_{p,y}$ denotes the return on asset Y. Once we compute σ_{xy} then we can easily trace out the efficient frontier.

To compute σ_{xy} we first note that

$$R_{p,x} = x_AR_A + x_BR_B + x_CR_C$$

and

$$R_{p,y} = y_AR_A + y_BR_B + y_CR_C.$$

Then, by the additivity of covariances, we have

$$\begin{aligned} \sigma_{xy} &= cov(x_AR_A + x_BR_B + x_CR_C, y_AR_A + y_BR_B + y_CR_C) && (9) \\ &= cov(x_AR_A, y_AR_A) + cov(x_AR_A, y_BR_B) + cov(x_AR_A, y_CR_C) \\ &\quad + cov(x_BR_B, y_AR_A) + cov(x_BR_B, y_BR_B) + cov(x_BR_B, y_CR_C) \\ &\quad + cov(x_CR_C, y_AR_A) + cov(x_CR_C, y_BR_B) + cov(x_CR_C, y_CR_C) \\ &= x_Ay_A\sigma^2_A + x_By_B\sigma^2_B + x_Cy_C\sigma^2_C \\ &\quad + (x_Ay_B + x_By_A)\sigma_{AB} + (x_Ay_C + x_Cy_A)\sigma_{AC} + (x_By_C + x_Cy_B)\sigma_{AC}. \end{aligned}$$

To illustrate these results, consider the example data with the previously computed efficient portfolios denoted asset X and asset Y. Consider forming a portfolio of these two portfolios with the weights $z_x = 0.5$ and $z_y = 0.5$. Then by straightforward calculations we have

$$\begin{aligned} \sigma_{xy} &= -1.163 \\ \mu_{p,z} &= (0.5)(0.01) + (0.5)(0.25) = 0.130 \\ \sigma^2_{p,z} &= (0.5)^2(1.066) + (0.5)^2(1.316) + 2(0.5)(0.5)(-1.163) = 0.014 \\ \sigma_{p,z} &= \sqrt{0.014} = 0.118 \end{aligned}$$

Portfolio Z is an efficient portfolio and the pair $(\mu_{p,z}, \sigma_{p,z}) = (0.130, 0.118)$ lies on the efficient frontier. This point is illustrated in figure xxx. To trace out the entire frontier we simply vary the weights z_x and z_y over some range, say $(z_x, z_y) = (0, 1)$, $(0.1, 0.9)$, . . . , $(1, 0)$,compute (7) and (8) and plot $\mu_{p,z}$ against $\sigma_{p,z}$. This is illustrated in figure xxx.

1.1 Finding the Global Minimum Variance Portfolio

The global minimum variance portfolio $\mathbf{m} = (m_A, m_B, m_C)'$ for the three asset case solves the constrained minimization problem

$$\begin{aligned} \min_{m_A,m_B,m_C} \sigma^2_{p,m} &= m_A^2\sigma_A^2 + m_B^2\sigma_B^2 + m_C^2\sigma_C^2 \\ &\quad +2m_Am_B\sigma_{AB} + 2m_Am_C\sigma_{AC} + 2m_Bm_C\sigma_{BC} \quad s.t. \\ 1 &= m_A + m_B + m_C. \end{aligned} \tag{10}$$

The Lagrangian for this problem is

$$\begin{aligned} L(m_A, m_B, m_C, \lambda) &= m_A^2\sigma_A^2 + m_B^2\sigma_B^2 + m_C^2\sigma_C^2 + 2m_Am_B\sigma_{AB} + 2m_Am_C\sigma_{AC} + 2m_Bm_C\sigma_{BC} \\ &\quad +\lambda(m_A + m_B + m_C - 1), \end{aligned}$$

and the first order conditions for a minimum are

$$\begin{aligned} 0 &= \frac{\partial L}{\partial m_A} = 2m_A\sigma_A^2 + 2m_B\sigma_{AB} + 2m_C\sigma_{AB} + \lambda \\ 0 &= \frac{\partial L}{\partial m_A} = 2m_B\sigma_B^2 + 2m_A\sigma_{AB} + 2m_C\sigma_{BC} + \lambda \\ 0 &= \frac{\partial L}{\partial m_A} = 2m_C\sigma_C^2 + 2m_A\sigma_{AC} + 2m_B\sigma_{BC} + \lambda \\ 0 &= \frac{\partial L}{\partial \lambda} = m_A + m_B + m_C - 1 \end{aligned} \tag{11}$$

This gives four linear equations in four unknowns which can be solved to find the global minimum variance portfolio.

Using the data in Table 1, it can be shown that the global minimum variance portfolio is $m_A = 0.310, m_B = 0.196$ and $m_C = 0.495$. The expected return, variance and standard deviation of this portfolio are

$$\begin{aligned} \mu_{p,m} &= \mu_{p,1} = (0.310)(0.229) + (0.196)(0.138) + (0.495)(0.528) \\ &= 0.124 \\ \sigma^2_{p,m} &= (0.310)^2(0.924) + (0.196)(0.862) + (0.495)(0.528) \\ &\quad +2(0.310)(0.196)(0.063) + 2(0.310)(0.495)(-0.582) + 2(0.196)(0.495)(-0.359) \\ &= 0.011 \\ \sigma_{p,m} &= \sqrt{0.011} = 0.103. \end{aligned}$$

The pair $(\mu_{p,m}, \sigma_{p,m}) = (0.011, 0.103)$ is illustrated in figure xxx.

1.2 Adding a Risk-Free Asset

Consider adding a risk-free asset (T-bill) with known return r_f to the investment problem. From our analysis of portfolios of two risky assets and a risk-free asset we know from the *mutual fund separation theorem* that the efficient set of portfolios are combinations of the risk-free asset and the so-called tangency portfolio. The tangency portfolio is the portfolio of risky assets that has the largest Sharpe's slope. Let t_A, t_B and t_C denote the proportions of assets A, B and C in the tangency portfolio. To find the tangency portfolio[2], we solve

$$\max_{t_A, t_B, t_C} \frac{\mu_{p,t} - r_f}{\sigma_{p,t}}$$

where

$$\begin{aligned}
\mu_{p,t} &= t_A\mu_A + t_B\mu_B + t_C\mu_C, \\
\sigma^2_{p,t} &= t_A^2\sigma_A^2 + t_B^2\sigma_B^2 + t_C^2\sigma_C^2 + 2t_At_B\sigma_{AB} + 2t_At_C\sigma_{AC} + 2t_Bt_C\sigma_{BC}.
\end{aligned}$$

Using the data from Table 1 and assuming a risk-free rate of $r_f = 0.12$, it can be shown that the tangency portfolio is $t_A = 0.532, t_B = 0.153$ and $t_C = 0.315$. The expected return, variance and standard deviation of this portfolio are

$$\begin{aligned}
\mu_{p,t} &= (0.532)(0.229) + (0.153)(0.138) + (0.315)(0.528) \\
&= 0.159 \\
\sigma^2_{p,t} &= (0.532)^2(0.924) + (0.153)(0.862) + (0.315)(0.528) \\
&\quad +2(0.532)(0.153)(0.063) + 2(0.532)(0.315)(-0.582) + 2(0.153)(0.315)(-0.359) \\
&= 0.115 \\
\sigma_{p,t} &= \sqrt{0.115} = 0.339.
\end{aligned}$$

The pair $(\mu_{p,t}, \sigma_{p,t}) = (0.115, 0.339)$ is illustrated in figure xxx.

The tangency portfolio can also be found analytically using the formula for the tangency portfolio in the case of two risky assets. In order to use this formula, however, the two risky assets must be efficient portfolios. To illustrate, consider the two efficient portfolios, portfolios X and Y, that solve (4) and (6). These portfolios have expected returns and variances $\mu_{p,x}, \mu_{p,y}, \sigma^2_{p,x}$ and $\sigma^2_{p,y}$. In addition, the covariance between the returns on these two portfolios is σ_{xy}. Let t_x denote the share of wealth in portfolio X and $t_y = 1 - t_x$ denote the share of wealth in portfolio Y. Then, using the analytic formula for the two risky asset case, we have

$$t_x = \frac{(\mu_{p,x} - r_f)\sigma^2_{p,y} - (\mu_{p,y} - r_f)\sigma_{xy}}{(\mu_{p,x} - r_f)\sigma^2_{p,y} + (\mu_{p,y} - r_f)\sigma^2_{p,x} - (\mu_{p,x} - r_f + \mu_{p,y} - r_f)\sigma_{xy}}, \; t_y = 1 - t_x. \quad (12)$$

[2]This is a very tedious calculus problem. However, it is easily solved numerically using the Solver in EXCEL.

The expected return and variance of this portfolio are

$$\begin{aligned}
\mu_{p,t} &= t_x\mu_{p,x} + t_y\mu_{p,y}, \\
\sigma^2_{p,t} &= {t_x}^2\sigma^2_{p,x} + t_y^2\sigma^2_{p,y} + 2t_xt_y\sigma_{xy}.
\end{aligned}$$

To illustrate this result using the data in Table 1, recall that $\mu_{p,x} = 0.01, \mu_{p,y} = 0.25, \sigma^2_{p,x} = 1.066$, $\sigma^2_{p,y} = 1.316$ and $\sigma_{xy} = -1.163$. Substituting these values into (12) gives $t_x = 0.378$ and $t_y = 0.622$. The expected return, variance and standard deviation of the tangency portfolio are

$$\begin{aligned}
\mu_{p,t} &= (0.378)(0.01) + (0.622)(0.25) = 0.159, \\
\sigma^2_{p,t} &= (0.378)^2(1.066) + (0.622)(1.316) + 2(0.378)(0.622)(-1.163) = 0.115, \\
\sigma_{p,t} &= 0.103,
\end{aligned}$$

which are the same as those found above. The weights in assets A,B and C in the tangency portfolio are

$$\begin{aligned}
t_A &= t_xx_A + t_yy_A = (0.378)(-0.398) + (0.622)(1.097) = 0.532 \\
t_B &= t_xx_B + t_yy_B = (0.378)(0.331) + (0.622)(0.045) = 0.153, \\
t_C &= t_xx_C + t_yy_C = (0.378)(1.067) + (0.622)(-0.142) = 0.315,
\end{aligned}$$

which are identical to those found above.

2 Portfolio math with matrix algebra

When working with large portfolios, the simple algebra of representing portfolio means and variances becomes cumbersome. The use of matrix (linear) algebra can greatly simplify many of the computations. Matrix algebra formulations are also very useful when it comes time to do actual computations on the computer. Popular spreadsheet programs like Excel and Lotus 123, which are the workhorse programs of many financial houses, can handle basic matrix calculations which also make it worthwhile to become familiar with matrix techniques[3].

Consider again the simple three asset portfolio problem. First, we define the following 3×1 column vectors containing the returns and portfolio weights

$$\mathbf{R} = \begin{pmatrix} R_A \\ R_B \\ R_C \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_A \\ x_B \\ x_C \end{pmatrix}.$$

In matrix notation we can lump multiple returns in a single vector which we denote by $\mathbf{R}$. Since each of the elements in $\mathbf{R}$ is a random variable we call $\mathbf{R}$ a *random vector.*

[3]The matrix functions available in Excel and Lotus 123 are very limited. Serious analysis should be done using matrix programming languages like Splus, Matlab or GAUSS.

We can also talk about the probability distribution of the random vector $\mathbf{R}$. This is simply the joint distribution of the elements of $\mathbf{R}$. In general, the distribution of $\mathbf{R}$ is complicated but if we assume that all returns are jointly normally distributed then all we need to worry about is the means, variances and covariances of the returns. We can easily summarize these values using matrix notation as follows. First, we define the 3×1 vector of portfolio expected values as

$$E[\mathbf{R}] = E\left[\begin{pmatrix} R_A \\ R_B \\ R_C \end{pmatrix}\right] = \begin{pmatrix} E[R_A] \\ E[R_B] \\ E[R_C] \end{pmatrix} = \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix} = \boldsymbol{\mu}$$

and the 3×3 covariance matrix of returns as

$$\begin{aligned} cov(\mathbf{R}) &= \begin{pmatrix} var(R_A) & cov(R_A, R_B) & cov(R_A, R_C) \\ cov(R_B, R_A) & var(R_B) & cov(R_B, R_C) \\ cov(R_C, R_A) & cov(R_C, R_B) & var(R_C) \end{pmatrix} \\ &= \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix} = \boldsymbol{\Sigma} \end{aligned}$$

Notice that the covariance matrix is symmetric (elements off the diagonal are equal so that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'$, where $\boldsymbol{\Sigma}'$ denotes the transpose of $\boldsymbol{\Sigma}$) since $cov(R_A, R_B) = cov(R_B, R_A)$, $cov(R_A, R_C) = cov(R_C, R_A)$ and $cov(R_B, R_C) = cov(R_C, R_B)$. Using the example data in Table 1 we have

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix} = \begin{pmatrix} 0.229 \\ 0.138 \\ 0.052 \end{pmatrix}, \\ \boldsymbol{\Sigma} &= \begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix}. \end{aligned}$$

The return on the portfolio using vector notation is

$$R_{p,x} = \mathbf{x}'\mathbf{R} = (x_A, x_B, x_C) \cdot \begin{pmatrix} R_A \\ R_B \\ R_C \end{pmatrix} = x_A R_A + x_B R_B + x_C R_C.$$

Similarly, the expected return on the portfolio is

$$\mu_{p,x} = E[\mathbf{x}'\mathbf{R}] = \mathbf{x}'E[\mathbf{R}] = \mathbf{x}'\boldsymbol{\mu} = (x_A, x_B, x_C) \cdot \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix} = x_A\mu_A + x_B\mu_B + x_C\mu_C.$$

Next, the variance of the portfolio is

$$\begin{aligned}
\sigma_{p,x}^2 &= var(\mathbf{x}'\mathbf{R}) = \mathbf{x}'\Sigma\mathbf{x} = (x_A, x_B, x_C) \cdot \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix} \begin{pmatrix} x_A \\ x_B \\ x_C \end{pmatrix} \\
&= x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC}
\end{aligned}$$

Finally, the condition that the portfolio weights sum to one can be expressed as

$$\mathbf{x}'\mathbf{1} = (x_A, x_B, x_C) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = x_A + x_B + x_C = 1$$

where $\mathbf{1}$ is a 3×1 vector with each element equal to 1.

Consider another portfolio with weights $\mathbf{y} = (y_A, y_B, y_C)'$. The return on this portfolio is

$$R_{p,y} = \mathbf{y}'\mathbf{R} = y_AR_A + y_BR_B + y_CR_C.$$

In the following we will need to compute the covariance between the return on portfolio $\mathbf{x}$ and the return on portfolio $\mathbf{y}$, $cov(R_{p,x}, R_{p,y})$. It can be shown that

$$\begin{aligned}
\sigma_{xy} &= cov(R_{p,x}, R_{p,y}) = cov(\mathbf{x}'\mathbf{R}, \mathbf{y}'\mathbf{R}) \\
&= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{y} = (x_A, x_B, x_C) \cdot \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix} \begin{pmatrix} y_A \\ y_B \\ y_C \end{pmatrix} \\
&= x_Ay_A\sigma_A^2 + x_By_B\sigma_B^2 + x_Cy_C\sigma_C^2 \\
&\quad +(x_Ay_B + x_By_A)\sigma_{AB} + (x_Ay_C + x_Cy_A)\sigma_{AC} + (x_By_C + x_Cy_B)\sigma_{AC}.
\end{aligned}$$

2.1 Finding Efficient Portfolios

The constrained minimization problem (4) to find an efficient portfolio can be re-expressed using matrix algebra as

$$\begin{aligned}
\min_{\mathbf{x}}\ \sigma_{p,x}^2 &= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}\ \ s.t. \\
\mu_{p,0} &= \mathbf{x}'\boldsymbol{\mu} \\
1 &= \mathbf{x}'\mathbf{1}
\end{aligned}$$

where $\mu_{p,0}$ is a target expected return. Matrix algebra can also be used to give an analytic solution to the first order conditions from the minimization problem (4). Since the first order conditions (5) consist of five linear equations in five unknowns $(x_A, x_B, x_C, \lambda_1, \lambda_2)$ we can represent the system in matrix notation as

$$\begin{pmatrix} 2\sigma_A^2 & 2\sigma_{AB} & 2\sigma_{AC} & \mu_A & 1 \\ 2\sigma_{AB} & 2\sigma_B^2 & 2\sigma_{BC} & \mu_B & 1 \\ 2\sigma_{AC} & 2\sigma_{BC} & 2\sigma_C^2 & \mu_C & 1 \\ \mu_A & \mu_B & \mu_C & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_A \\ x_B \\ x_C \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu_{p,0} \\ 1 \end{pmatrix}$$

or

$$\mathbf{A}\mathbf{z}_x = \mathbf{b}_0$$

where

$$\mathbf{A} = \begin{pmatrix} 2\sigma_A^2 & 2\sigma_{AB} & 2\sigma_{AC} & \mu_A & 1 \\ 2\sigma_{AB} & 2\sigma_B^2 & 2\sigma_{BC} & \mu_B & 1 \\ 2\sigma_{AC} & 2\sigma_{BC} & 2\sigma_C^2 & \mu_C & 1 \\ \mu_A & \mu_B & \mu_C & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}, \ \mathbf{z}_x = \begin{pmatrix} x_A \\ x_B \\ x_C \\ \lambda_1 \\ \lambda_2 \end{pmatrix} \text{ and } \mathbf{b}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu_{p,0} \\ 1 \end{pmatrix}$$

The solution for $\mathbf{z}_x$ is then

$$\mathbf{z}_x = \mathbf{A}^{-1}\mathbf{b}_0.$$

The first three elements of $\mathbf{z}_x$ are the portfolio weights $\mathbf{x} = (x_A, x_B, x_C)'$ for the efficient portfolio with expected return $\mu_{p,x} = \mu_{p,0}$ and standard deviation $\sigma_{p,x}$.

To illustrate consider the data in Table 1 and the target expected return $\mu_{p,0} = 0.01$. Then

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} 1.848 & 0.126 & -1.164 & 0.229 & 1 \\ 0.126 & 1.724 & -0.718 & 0.138 & 1 \\ -1.164 & -0.718 & 1.056 & 0.052 & 1 \\ 0.229 & 0.138 & 0.057 & 0.052 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}, \\ \mathbf{A}^{-1} &= \begin{pmatrix} 0.095 & -0.196 & 0.101 & 6.229 & -0.460 \\ -0.196 & 0.404 & -0.208 & -1.192 & 0.343 \\ 0.101 & -0.208 & 0.107 & -5.037 & 1.117 \\ 6.229 & -1.192 & -5.037 & -163.5 & 20.22 \\ -0.460 & 0.343 & 1.117 & 20.22 & -2.521 \end{pmatrix}, \ \mathbf{b}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.01 \\ 1 \end{pmatrix} \end{aligned}$$

and

$$\mathbf{z}_x = \begin{pmatrix} -0.398 \\ 0.331 \\ 1.067 \\ 18.58 \\ -2.319 \end{pmatrix} = \begin{pmatrix} 0.095 & -0.196 & 0.101 & 6.229 & -0.460 \\ -0.196 & 0.404 & -0.208 & -1.192 & 0.343 \\ 0.101 & -0.208 & 0.107 & -5.037 & 1.117 \\ 6.229 & -1.192 & -5.037 & -163.5 & 20.22 \\ -0.460 & 0.343 & 1.117 & 20.22 & -2.521 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.01 \\ 1 \end{pmatrix}$$

Hence, the efficient portfolio is $\mathbf{x} = (-0.398, 0.331, 1.067)'$. The expected return on this portfolio is

$$\begin{aligned} \mu_{p,x} &= \mathbf{x}'\boldsymbol{\mu} \\ &= (-0.398, 0.331, 1.067) \cdot \begin{pmatrix} 0.229 \\ 0.138 \\ 0.052 \end{pmatrix} = 0.01 \end{aligned}$$

and the variance is

$$\begin{aligned}
\sigma^2_{p,x} &= \mathbf{x}'\mathbf{\Sigma}\mathbf{x} \\
&= (-0.398, 0.331, 1.067) \cdot \begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix} \begin{pmatrix} -0.398 \\ 0.331 \\ 1.067 \end{pmatrix} = 1.066.
\end{aligned}$$

To find another efficient portfolio $\mathbf{y} = (y_A, y_B, y_C)'$ we solve

$$\begin{aligned}
\min_{\mathbf{y}} \ \sigma^2_{p,y} &= \mathbf{y}'\mathbf{\Sigma}\mathbf{y} \quad s.t. \\
\mu_{p,1} &= \mathbf{y}'\boldsymbol{\mu} \\
1 &= \mathbf{y}'\mathbf{1}
\end{aligned}$$

where $\mu_{p,1}$ is a target expected return different from $\mu_{p,0}$. The solution has the form

$$\mathbf{A}\mathbf{z}_y = \mathbf{b}_1$$

with

$$\mathbf{z}_y = \begin{pmatrix} y_A \\ y_B \\ y_C \\ \lambda_1 \\ \lambda_2 \end{pmatrix} \text{ and } \mathbf{b}_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \mu_{p,1} \\ 1 \end{pmatrix}.$$

The first three elements of $\mathbf{z}_y$ are the portfolio weights $\mathbf{y} = (y_A, y_B, y_C)'$ for the efficient portfolio with expected return $\mu_{p,y} = \mu_{p,1}$ and standard deviation $\sigma_{p,y}$.

Using the data in table 1 with the target expected return $\mu_{p,1} = 0.25$ we have

$$\mathbf{z}_y = \begin{pmatrix} 1.097 \\ 0.045 \\ -0.142 \\ 20.66 \\ 2.533 \end{pmatrix} = \begin{pmatrix} 0.095 & -0.196 & 0.101 & 6.229 & -0.460 \\ -0.196 & 0.404 & -0.208 & -1.192 & 0.343 \\ 0.101 & -0.208 & 0.107 & -5.037 & 1.117 \\ 6.229 & -1.192 & -5.037 & -163.5 & 20.22 \\ -0.460 & 0.343 & 1.117 & 20.22 & -2.521 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.25 \\ 1 \end{pmatrix}$$

Hence, the second efficient portfolio is $\mathbf{y} = (1.097, 0.045, -0.142)'$. The expected return on this portfolio is

$$\begin{aligned}
\mu_{p,y} &= \mathbf{y}'\boldsymbol{\mu} \\
&= (1.097, 0.045, -0.142) \cdot \begin{pmatrix} 0.229 \\ 0.138 \\ 0.052 \end{pmatrix} = 0.25
\end{aligned}$$

and the variance is

$$\begin{aligned}
\sigma^2_{p,y} &= \mathbf{y}'\mathbf{\Sigma}\mathbf{y} \\
&= (1.097, 0.045, -0.142) \cdot \begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix} \begin{pmatrix} 1.097 \\ 0.045 \\ -0.142 \end{pmatrix} = 1.316.
\end{aligned}$$

2.2 Finding the Global Minimum Variance Portfolio

Using matrix notation, the problem (10) may be concisely expressed as

$$\begin{aligned} \min_{\mathbf{m}} \sigma^2_{p,m} &= \mathbf{m}'\mathbf{\Sigma}\mathbf{m} \ s.t. \\ 1 &= \mathbf{m}'\mathbf{1}. \end{aligned}$$

The four linear equation describing the first order conditions (11) has the matrix representation

$$\begin{pmatrix} 2\sigma_A^2 & 2\sigma_{AB} & 2\sigma_{AC} & 1 \\ 2\sigma_{AB} & 2\sigma_B^2 & 2\sigma_{BC} & 1 \\ 2\sigma_{AC} & 2\sigma_{BC} & 2\sigma_C^2 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} m_A \\ m_B \\ m_C \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

or

$$\mathbf{C}\mathbf{z}_m = \mathbf{b}$$

where

$$\mathbf{C} = \begin{pmatrix} 2\sigma_A^2 & 2\sigma_{AB} & 2\sigma_{AC} & 1 \\ 2\sigma_{AB} & 2\sigma_B^2 & 2\sigma_{BC} & 1 \\ 2\sigma_{AC} & 2\sigma_{BC} & 2\sigma_C^2 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \ \mathbf{z}_m = \begin{pmatrix} m_A \\ m_B \\ m_C \\ \lambda \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

The solution for $\mathbf{z}_m$ is then

$$\mathbf{z}_m = \mathbf{C}^{-1}\mathbf{b}.$$

The first three elements of $\mathbf{z}_m$ are the portfolio weights $\mathbf{m} = (m_A, m_B, m_C)'$ for the global minimum variance portfolio with expected return $\mu_{p,m} = \mathbf{m}'\boldsymbol{\mu}$ and variance $\sigma^2_{p,m} = \mathbf{m}'\mathbf{\Sigma}\mathbf{m}$.

Using the data in Table 1, we have

$$\begin{aligned} C &= \begin{pmatrix} 1.848 & 0.126 & -1.164 & 1 \\ 0.126 & 1.724 & -0.718 & 1 \\ -1.164 & -0.718 & 1.056 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \\ C^{-1} &= \begin{pmatrix} 0.333 & -0.242 & -0.091 & 0.310 \\ -0.242 & 0.413 & -0.171 & 0.196 \\ -0.091 & -0.171 & 0.262 & 0.495 \\ 0.310 & 0.196 & 0.495 & -0.021 \end{pmatrix} \end{aligned}$$

and so

$$\mathbf{z}_m = \begin{pmatrix} 0.333 & -0.242 & -0.091 & 0.310 \\ -0.242 & 0.413 & -0.171 & 0.196 \\ -0.091 & -0.171 & 0.262 & 0.495 \\ 0.310 & 0.196 & 0.495 & -0.021 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.310 \\ 0.196 \\ 0.495 \\ -0.021 \end{pmatrix}.$$

Hence, the global minimum variance portfolio is $\mathbf{m} = (0.310, 0.196, 0.495)'$. The expected return on this portfolio is

$$
\begin{aligned}
\mu_{p,m} &= \mathbf{m}'\boldsymbol{\mu} \\
&= (0.310, 0.196, 0.495)\begin{pmatrix} 0.229 \\ 0.138 \\ 0.052 \end{pmatrix} = 0.124
\end{aligned}
$$

and the variance is

$$
\begin{aligned}
\sigma^2_{p,m} &= \mathbf{m}'\boldsymbol{\Sigma}\mathbf{m} \\
&= (0.310, 0.196, 0.495)\cdot\begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix}\begin{pmatrix} 0.310 \\ 0.196 \\ 0.495 \end{pmatrix} = 0.011.
\end{aligned}
$$

2.3 Computing the Efficient Frontier

As mentioned previously, to compute the efficient frontier or Markowitz bullet one only needs to find two efficient portfolios. The remaining efficient portfolios can then be expressed as convex combinations of these two portfolios. The following proposition describes the process for the three risky asset case using matrix algebra.

Proposition 1 *Let* $\mathbf{x} = (x_A, x_B, x_C)'$ *and* $\mathbf{y} = (y_A, y_B, y_C)'$ *be any two efficient portfolios. That is,* $\mathbf{x}$ *solves*

$$
\begin{aligned}
\min_{\mathbf{x}}\ \sigma^2_{p,x} &= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}\ \ s.t. \\
\mu_{p,0} &= \mathbf{x}'\boldsymbol{\mu} \\
1 &= \mathbf{x}'\mathbf{1}
\end{aligned}
$$

and $\mathbf{y}$ *solves*

$$
\begin{aligned}
\min_{\mathbf{y}}\ \sigma^2_{p,y} &= \mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}\ \ s.t. \\
\mu_{p,1} &= \mathbf{y}'\boldsymbol{\mu} \\
1 &= \mathbf{y}'\mathbf{1}
\end{aligned}
$$

Let α *be any constant. Then the portfolio*

$$
\begin{aligned}
\mathbf{z} &= \alpha\cdot\mathbf{x} + (1-\alpha)\cdot\mathbf{y} \\
&= \begin{pmatrix} \alpha x_A + (1-\alpha)y_A \\ \alpha x_B + (1-\alpha)y_B \\ \alpha x_C + (1-\alpha)y_C \end{pmatrix}
\end{aligned}
$$

is an efficient portfolio. Furthermore,

$$\begin{aligned}
\mu_{p,z} &= \mathbf{z}'\boldsymbol{\mu} = \alpha \cdot \mu_{p,x} + (1-\alpha) \cdot \mu_{p,y} \\
\sigma^2_{p,z} &= \mathbf{z}'\boldsymbol{\Sigma}\mathbf{z} = \boldsymbol{\alpha}^2\sigma^2_{p,x} + (1-\alpha)^2\sigma^2_{p,y} + 2\alpha(1-\alpha)\sigma_{xy}
\end{aligned}$$

where

$$\begin{aligned}
\sigma^2_{p,x} &= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}, \\
\sigma^2_{p,y} &= \mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}, \\
\sigma_{xy} &= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{y}.
\end{aligned}$$

To illustrate the practical application of the proposition, we will use the data in Table 1 and the previously computed efficient portfolios $\mathbf{x} = (-0.398, 0.331, 1.067)'$ and $\mathbf{y} = (1.097, 0.045, -0.142)'$. Recall, that $\mu_{p,x} = 0.01, \sigma^2_{p,x} = 1.066, \mu_{p,y} = 0.25$ and $\sigma^2_{p,y} = 1.316$. First, we need to compute the covariance between the return on portfolio $\mathbf{x}$ and the return on portfolio $\mathbf{y}$:

$$\begin{aligned}
\sigma_{xy} &= \mathbf{x}'\boldsymbol{\Sigma}\mathbf{y} \\
&= (-0.398, 0.331, 1.067) \cdot \begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix} \begin{pmatrix} 1.097 \\ 0.045 \\ -0.142 \end{pmatrix} = -1.163.
\end{aligned}$$

Next, consider convex combinations of $\mathbf{x}$ and $\mathbf{y}$ with the constant α ranging from 0 to 1 in increments of 0.1. For example, when $\alpha = 0.5$ the portfolio $\mathbf{z}$ becomes

$$\begin{aligned}
\mathbf{z} &= \alpha \cdot \mathbf{x} + (1-\alpha) \cdot \mathbf{y} \\
&= 0.5 \cdot \begin{pmatrix} -0.398 \\ 0.331 \\ 1.067 \end{pmatrix} + 0.5 \cdot \begin{pmatrix} 1.097 \\ 0.045 \\ -0.142 \end{pmatrix} \\
&= \begin{pmatrix} (0.5)(-0.398) \\ (0.5)(0.331) \\ (0.5)(1.067) \end{pmatrix} + \begin{pmatrix} (0.5)(-0.398) \\ (0.5)(0.331) \\ (0.5)(1.067) \end{pmatrix} \\
&= \begin{pmatrix} 0.349 \\ 0.188 \\ 0.463 \end{pmatrix} = \begin{pmatrix} z_A \\ z_B \\ z_C \end{pmatrix}.
\end{aligned}$$

The expected return and variance of this portfolio is

$$\begin{aligned}
\mu_{p,z} &= \mathbf{z}'\boldsymbol{\mu} \\
&= (0.349, 0.188, 0.463) \cdot \begin{pmatrix} 0.229 \\ 0.138 \\ 0.052 \end{pmatrix} = 0.130, \\
\sigma^2_{p,z} &= \mathbf{z}'\boldsymbol{\Sigma}\mathbf{z} \\
&= (0.349, 0.188, 0.463) \cdot \begin{pmatrix} 0.924 & 0.063 & -0.582 \\ 0.063 & 0.862 & -0.359 \\ -0.582 & -0.359 & 0.528 \end{pmatrix} \begin{pmatrix} 0.349 \\ 0.188 \\ 0.463 \end{pmatrix} = 0.014.
\end{aligned}$$

Note that $\mu_{p,z}$ and $\sigma_{p,z}^2$ can also be computed as

$$
\begin{aligned}
\mu_{p,z} &= \alpha\mu_{p,x} + (1-\alpha)\mu_{p,y} \\
&= (0.5)(0.01) + (0.5)(0.25) = 0.13, \\
\sigma_{p,z}^2 &= \boldsymbol{\alpha}^2\sigma_{p,x}^2 + (1-\alpha)^2\sigma_{p,y}^2 + 2\alpha(1-\alpha)\sigma_{xy} \\
&= (0.5)^2(1.066) + (0.5)^2(1.316) + 2(0.5)(0.5)(-1.163) = 0.014.
\end{aligned}
$$

The graph of $\mu_{p,z}$ against $\sigma_{p,z}$ for $\alpha \in (0,1)$ is exactly the same as that calculated in section xxx and is illustrated in figure xxx.

2.4 Computing the Tangency Portfolio

The tangency portfolio solves

$$\max_{\mathbf{t}} \frac{\mathbf{t}'\boldsymbol{\mu} - r_f}{(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})^{\frac{1}{2}}}$$

Alternatively, we can use (12) with two efficient portfolios $\mathbf{x}$ and y that solve (4) and (6).

3 Efficient Portfolios with N Risky Assets and a Risk free Asset Using Matrix Algebra

To be completed

4 Estimating the Inputs to the General Portfolio Problem

To be completed

4.1 Application: Global asset allocation

To be completed

5 Appendix A Digression on the Covariance Matrix

The covariance matrix of returns, $\boldsymbol{\Sigma}$, summarizes the variances and covariances of the individual returns in the return vector $\mathbf{R}$. In general, the covariance matrix of a random vector $\mathbf{R}$ (sometimes simply called the variance of vector $\mathbf{R}$) with mean vector $\boldsymbol{\mu}$ is defined as

$$cov(\mathbf{R}) = E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})'] = \mathbf{\Sigma}$$

If $\mathbf{R}$ has N elements then $\mathbf{\Sigma}$ will be an $N \times N$ matrix. For the case $N = 2$, we have

$$\begin{aligned}
E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})'] &= E\left[\begin{pmatrix} R_A-\mu_A \\ R_B-\mu_B \end{pmatrix} \cdot (R_A-\mu_A, R_B-\mu_B)\right] \\
&= E\left[\begin{pmatrix} (R_A-\mu_A)^2 & (R_A-\mu_A)(R_B-\mu_B) \\ (R_B-\mu_B)(R_A-\mu_A) & (R_B-\mu_B)^2 \end{pmatrix}\right] \\
&= \begin{pmatrix} E[(R_A-\mu_A)^2] & E[(R_A-\mu_A)(R_B-\mu_B)] \\ E[(R_B-\mu_B)(R_A-\mu_A)] & E[(R_B-\mu_B)^2] \end{pmatrix} \\
&= \begin{pmatrix} var(R_A) & cov(R_A,R_B) \\ cov(R_B,R_A) & var(R_B) \end{pmatrix} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} \\ \sigma_{AB} & \sigma_B^2 \end{pmatrix} = \mathbf{\Sigma}
\end{aligned}$$

We can use the formal definition of $cov(\mathbf{R})$ to derive the variance of a portfolio. Consider again the two asset case. The variance of the portfolio $R_p = \mathbf{x}'\mathbf{R}$ is given by

$$var(R_p) = var(\mathbf{x}'\mathbf{R}) = E[(\mathbf{x}'\mathbf{R}-\mathbf{x}'\boldsymbol{\mu})^2] = E[(\mathbf{x}'(\mathbf{R}-\boldsymbol{\mu}))^2]$$

since $\mathbf{x}'\mathbf{R}$ is a scalar. Now we use a trick from matrix algebra. If z is a scalar (think of $z = 2$) then $z'z = z \cdot z' = z^2$. Let $z = \mathbf{x}'(\mathbf{R}-\boldsymbol{\mu})$ and so $z \cdot z' = \mathbf{x}'(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})'\mathbf{x}$. Then

$$\begin{aligned}
var(R_p) &= E[z^2] = E[z \cdot z'] \\
&= E[\mathbf{x}'(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})'\mathbf{x}] \\
&= \mathbf{x}'E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})']\mathbf{x} \\
&= \mathbf{x}'cov(\mathbf{R})\mathbf{x} = \mathbf{x}'\mathbf{\Sigma}\mathbf{x}.
\end{aligned}$$

Next consider determining the covariance between the returns on two portfolios $\mathbf{x}$ and $\mathbf{y}$. The returns on these two portfolios are $R_{p,x} = \mathbf{x}'\mathbf{R}$ and $R_{p,y} = \mathbf{y}'\mathbf{R}$. From the definition of covariance we have

$$cov(R_{p,x}, R_{p,y}) = E[(R_{p,x}-\mu_{p,x})(R_{p,y}-\mu_{p,y})]$$

which may be rewritten in matrix notation as

$$\begin{aligned}
cov(\mathbf{x}'\mathbf{R}, \mathbf{y}'\mathbf{R}) &= E[(\mathbf{x}'\mathbf{R}-\mathbf{x}'\boldsymbol{\mu})(\mathbf{y}'\mathbf{R}-\mathbf{y}'\boldsymbol{\mu})] \\
&= E[\mathbf{x}'(\mathbf{R}-\boldsymbol{\mu})\mathbf{y}'(\mathbf{R}-\boldsymbol{\mu})] \\
&= E[\mathbf{x}'(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})'\mathbf{y}] \\
&= \mathbf{x}'E[(\mathbf{R}-\boldsymbol{\mu})(\mathbf{R}-\boldsymbol{\mu})']\mathbf{y} \\
&= \mathbf{x}'\mathbf{\Sigma}\mathbf{y}.
\end{aligned}$$

6 Problems

7 References

Introduction to Financial Econometrics Chapter 6 The Single Index Model and Bivariate Regression

Eric Zivot University of Washington
Department of Economics

March 1, 2001

1 The single index model

Sharpe's *single index model*, also know as the *market model* and the *single factor model*, is a purely statistical model used to explain the behavior of asset returns. It is a generalization of the *constant expected return* (CER) model to account for systematic factors that may affect an asset's return. It is not the same model as the *Capital Asset Pricing Model* (CAPM), which is an economic model of equilibrium returns, but is closely related to it as we shall see in the next chapter.

The single index model has the form of a simple bivariate linear regression model

$$R_{it} = \alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it},\ i = 1, \ldots, N; t = 1, \ldots, T \tag{1}$$

where R_{it} is the continuously compounded return on asset i ($i = 1, \ldots, N$) between time periods $t-1$ and t, and R_{Mt} is the continuously compounded return on a *market index* portfolio between time periods $t-1$ and t. The market index portfolio is usually some well diversified portfolio like the S&P 500 index, the Wilshire 5000 index or the CRSP[1] equally or value weighted index. As we shall see, the coefficient $\beta_{i,M}$ multiplying R_{Mt} in (1) measures the contribution of asset i to the variance (risk), σ_M^2, of the market index portfolio. If $\beta_{i,M} = 1$ then adding the security does not change the variability, σ_M^2, of the market index; if $\beta_{i,M} > 1$ then adding the security will increase the variability of the market index and if $\beta_{i,M} < 1$ then adding the security will decrease the variability of the market index.

The intuition behind the single index model is as follows. The market index R_{Mt} captures "macro" or market-wide systematic risk factors that affect all returns in one way or another. This type of risk, also called *covariance risk, systematic risk* and

[1] CRSP refers to the Center for Research in Security Prices at the University of Chicago.

market risk, cannot be eliminated in a well diversified portfolio. The random error term ε_{it} has a similar interpretation as the error term in the CER model. In the single index model, ε_{it} represents random "news" that arrives between time $t-1$ and t that captures "micro" or firm-specific risk factors that affect an individual asset's return that are not related to macro events. For example, ε_{it} may capture the news effects of new product discoveries or the death of a CEO. This type of risk is often called *firm specific risk*, *idiosyncratic risk*, *residual risk* or *non-market risk*. This type of risk can be eliminated in a well diversified portfolio.

The single index model can be expanded to capture multiple factors. The single index model then takes the form a $k-$variable linear regression model

$$R_{it} = \alpha_i + \beta_{i,1}F_{1t} + \beta_{i,2}F_{2t} + \cdots + \beta_{i,k}F_{kt} + \varepsilon_{it}$$

where F_{jt} denotes the j^{th} systematic factorm, $\beta_{i,j}$ denotes asset $i's$ loading on the j^{th} factor and ε_{it} denotes the random component independent of all of the systematic factors. The single index model results when $F_{1t} = R_{Mt}$ and $\beta_{i,2} = \cdots = \beta_{i,k} = 0$. In the literature on multiple factor models the factors are usually variables that capture specific characteristics of the economy that are thought to affect returns - e.g. the market index, GDP growth, unexpected inflation etc., and firm specific or industry specific characteristics - firm size, liquidity, industry concentration etc. Multiple factor models will be discussed in chapter xxx.

The single index model is heavily used in empirical finance. It is used to estimate expected returns, variances and covariances that are needed to implement portfolio theory. It is used as a model to explain the "normal" or usual rate of return on an asset for use in so-called *event studies*[2]. Finally, the single index model is often used the evaluate the performance of mutual fund and pension fund managers.

1.1 Statistical Properties of Asset Returns in the single index model

The statistical assumptions underlying the single index model (1) are as follows:

1. (R_{it}, R_{Mt}) are jointly normally distributed for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.
2. $E[\varepsilon_{it}] = 0$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$ (news is neutral on average).
3. $var(\varepsilon_{it}) = \sigma^2_{\varepsilon,i}$ for $i = 1, \ldots, N$ (homoskedasticity).
4. $cov(\varepsilon_{it}, R_{Mt}) = 0$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

[2] The purpose of an event study is to measure the effect of an economic event on the value of a firm. Examples of event studies include the analysis of mergers and acquisitions, earning announcements, announcements of macroeconomic variables, effects of regulatory change and damage assessments in liability cases. An excellent overview of event studies is given in chapter 4 of Campbell, Lo and MacKinlay (1997).

5. $cov(\varepsilon_{it}, \varepsilon_{js}) = 0$ for all t, s and $i \neq j$

6. ε_{it} is normally distributed

The normality assumption is justi&ed on the observation that returns are fairly well characterized by the normal distribution. The error term having mean zero implies that &rm speci&c news is, on average, neutral and the constant variance assumptions implies that the magnitude of typical news events is constant over time. Assumption 4 states that &rm speci&c news is independent (since the random variables are normally distributed) of macro news and assumption 5 states that news affecting asset i in time t is independent of news affecting asset j in time s.

That ε_{it} is unrelated to R_{Ms} and ε_{js} implies that any correlation between asset i and asset j is solely due to their common exposure to R_{Mt} throught the values of β_i and β_j.

1.1.1 Unconditional Properties of Returns in the single index model

The unconditional properties of returns in the single index model are based on the marginal distribution of returns: that is, the distribution of R_{it} without regard to any information about R_{Mt}. These properties are summarized in the following proposition.

Proposition 1 *Under assumptions 1 - 6*

1. $E[R_{it}] = \mu_i = \alpha_i + \beta_{i,M} E[R_{Mt}] = \alpha_i + \beta_{i,M}\mu_M$
2. $var(R_{it}) = \sigma_i^2 = \beta_{i,M}^2 var(R_{Mt}) + var(\varepsilon_{it}) = \beta_{i,M}^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$
3. $cov(R_{it}, R_{jt}) = \sigma_{ij} = \sigma_M^2 \beta_i \beta_j$
4. $R_{it} \sim$ iid $N(\mu_i, \sigma_i^2)$, $R_{Mt} \sim$ iid $N(\mu_M, \sigma_M^2)$
5. $\beta_{i,M} = \frac{cov(R_{it}, R_{Mt})}{var(R_{Mt})} = \frac{\sigma_{iM}}{\sigma_M^2}$

The proofs of these results are straightforward and utilize the properties of linear combinations of random variables. Results 1 and 4 are trivial. For 2, note that

$$\begin{aligned} var(R_{it}) &= var(\alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it}) \\ &= \beta_{i,M}^2 var(R_{Mt}) + var(\varepsilon_{it}) + 2cov(R_{Mt}, \varepsilon_{it}) \\ &= \beta_{i,M}^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2 \end{aligned}$$

since, by assumption 4, $cov(\varepsilon_{it}, R_{Mt}) = 0$. For 3, by the additivity property of covariance and assumptions 4 and 5 we have

$$\begin{aligned} cov(R_{it}, R_{jt}) &= cov(\alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it}, \alpha_j + \beta_{j,M} R_{Mt} + \varepsilon_{jt}) \\ &= cov(\beta_{i,M} R_{Mt} + \varepsilon_{it}, \beta_{j,M} R_{Mt} + \varepsilon_{jt}) \\ &= cov(\beta_{i,M} R_{Mt}, \beta_{j,M} R_{Mt}) + cov(\beta_{i,M} R_{Mt}, \varepsilon_{jt}) + cov(\varepsilon_{it}, \beta_{j,M} R_{Mt}) + cov(\varepsilon_{it}, \varepsilon_{jt}) \\ &= \beta_{i,M} \beta_{j,M} cov(R_{Mt}, R_{Mt}) = \beta_{i,M} \beta_{j,M} \sigma_M^2 \end{aligned}$$

Last, for 5 note that

$$\begin{aligned} cov(R_{it}, R_{Mt}) &= cov(\alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it}, R_{Mt}) \\ &= cov(\beta_{i,M} R_{Mt}, R_{Mt}) \\ &= \beta_{i,M} cov(R_{Mt}, R_{Mt}) \\ &= \beta_{i,M} var(R_{Mt}), \end{aligned}$$

which uses assumption 4. It follows that

$$\frac{cov(R_{it}, R_{Mt})}{var(R_{Mt})} = \frac{\beta_{i,M} var(R_{Mt})}{var(R_{Mt})} = \beta_{i,M}.$$

Remarks:

1. Notice that unconditional expected return on asset i, μ_i, is constant and consists of an intercept term α_i, a term related to $\beta_{i,M}$ and the unconditional mean of the market index, μ_M. This relationship may be used to create predictions of expected returns over some future period. For example, suppose $\alpha_i = 0.01, \beta_{i,M} = 0.5$ and that a market analyst forecasts $\mu_M = 0.05$. Then the forecast for the expected return on asset i is

$$\widehat{\mu}_i = 0.01 + 0.5(0.05) = 0.026.$$

2. The unconditional variance of the return on asset i is constant and consists of variability due to the market index, $\beta_{i,M}^2 \sigma_M^2$, and variability due to speci&c risk, $\sigma_{\varepsilon,i}^2$.

3. Since $\sigma_{ij} = \sigma_M^2 \beta_i \beta_j$ the direction of the covariance between asset i and asset j depends of the values of β_i and β_j. In particular

   - $\sigma_{ij} = 0$ if $\beta_i = 0$ or $\beta_j = 0$ or both
   - $\sigma_{ij} > 0$ if β_i and β_j are of the same sign
   - $\sigma_{ij} < 0$ if β_i and β_j are of opposite signs.

4. The expression for the expected return can be used to provide an unconditional interpretation of α_i. Subtracting $\beta_{i,M}\mu_M$ from both sides of the expression for μ_i gives

$$\alpha_i = \mu_i - \beta_{i,M}\mu_M.$$

1.1.2 Decomposing Total Risk

The independence assumption between R_{Mt} and ε_{it} allows the unconditional variability of $R_{it}, var(R_{it}) = \sigma_i^2$, to be decomposed into the variability due to the market index, $\beta_{i,M}^2\sigma_M^2$, plus the variability of the &rm speci&c component, $\sigma_{\varepsilon,i}^2$. This decomposition is often called analysis of variance (ANOVA). Given the ANOVA, it is useful to de&ne the *proportion* of the variability of asset i that is due to the market index and the *proportion* that is unrelated to the index. To determine these proportions, divide both sides of $\sigma_i^2 = \beta_{i,M}^2\sigma_M^2 + \sigma_{\varepsilon,i}^2$ to give

$$1 = \frac{\sigma_i^2}{\sigma_i^2} = \frac{\beta_{i,M}^2\sigma_M^2 + \sigma_{\varepsilon,i}^2}{\sigma_i^2} = \frac{\beta_{i,M}^2\sigma_M^2}{\sigma_i^2} + \frac{\sigma_{\varepsilon,i}^2}{\sigma_i^2}$$

Then we can de&ne

$$R_i^2 = \frac{\beta_{i,M}^2\sigma_M^2}{\sigma_i^2} = 1 - \frac{\sigma_{\varepsilon,i}^2}{\sigma_i^2}$$

as the proportion of the total variability of R_{it} that is attributable to variability in the market index. Similarly,

$$1 - R_i^2 = \frac{\sigma_{\varepsilon,i}^2}{\sigma_i^2}$$

is then the proportion of the variability of R_{it} that is due to &rm speci&c factors. We can think of R_i^2 as measuring the proportion of risk in asset i that cannot be diversi&ed away when forming a portfolio and we can think of $1-R_i^2$ as the proportion of risk that can be diversi&ed away. It is important not to confuse R_i^2 with $\beta_{i,M}$. The coefficient $\beta_{i,M}$ measures the overall magnitude of nondiversi&able risk whereas R_i^2 measures the proportion of this risk in the total risk of the asset.

William Sharpe computed R_i^2 for thousands of assets and found that for a typical stock $R_i^2 \approx 0.30$. That is, 30% of the variability of the return on a typical is due to variability in the overall market and 70% of the variability is due to non-market factors.

1.1.3 Conditional Properties of Returns in the single index model

Here we refer to the properties of returns conditional on observing the value of the market index random variable R_{Mt}. That is, suppose it is known that $R_{Mt} = r_{Mt}$. The following proposition summarizes the properties of the single index model conditional on $R_{Mt} = r_{Mt}$:

1. $E[R_{it}|R_{Mt} = r_{Mt}] = \mu_{i|R_M} = \alpha_i + \beta_{i,M}r_{Mt}$
2. $var(R_{it}|R_{Mt} = r_{Mt}) = var(\varepsilon_{it}) = \sigma_{\varepsilon,i}^2$
3. $cov(R_{it}, R_{jt}|R_{mt} = r_{Mt}) = 0$

Property 1 states that the expected return on asset i conditional on $R_{Mt} = r_{Mt}$ is allowed to vary with the level of the market index. Property 2 says conditional on the value of the market index, the variance of the return on asset is equal to the variance of the random news component. Property 3 shows that once movements in the market are controlled for, assets are uncorrelated.

1.2 Matrix Algebra Representation of the Single Index Model

The single index model for the entire set of N assets may be conveniently represented using matrix algebra. De&nie the $(N \times 1)$ vectors $\mathbf{R}_t = (R_{1t}, R_{2t}, \ldots, R_{Nt})'$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. Then the single index model for all N assets may be represented as

$$\begin{pmatrix} R_{1t} \\ \vdots \\ R_{Nt} \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix} + \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} R_{Mt} + \begin{pmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{pmatrix},\ t = 1, \ldots, T$$

or

$$\mathbf{R}_t = \boldsymbol{\alpha} + \boldsymbol{\beta} \cdot R_{Mt} + \boldsymbol{\varepsilon}_t,\ t = 1, \ldots, T.$$

Since $\sigma_i^2 = \beta_{i,M}^2 \sigma_M^2 + \sigma_{\varepsilon,i}^2$ and $\sigma_{ij} = \beta_i \beta_j \sigma_M^2$ the covariance matrix for the N returns may be expressed as

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \cdots & \cdots & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} \beta_{i,M}^2 \sigma_M^2 & \beta_i \beta_j \sigma_M^2 & \cdots & \beta_i \beta_j \sigma_M^2 \\ \beta_i \beta_j \sigma_M^2 & \beta_{i,M}^2 \sigma_M^2 & \cdots & \beta_i \beta_j \sigma_M^2 \\ \vdots & \vdots & \ddots & \vdots \\ \beta_i \beta_j \sigma_M^2 & \beta_i \beta_j \sigma_M^2 & \cdots & \beta_{i,M}^2 \sigma_M^2 \end{pmatrix} + \begin{pmatrix} \sigma_{\varepsilon,1}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{\varepsilon,2}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{\varepsilon,N}^2 \end{pmatrix}$$

The covariance matrix may be conveniently computes as

$$\boldsymbol{\Sigma} = \sigma_M^2 \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Delta}$$

where Δ is a diagonal matrix with $\sigma_{\varepsilon,i}^2$ along the diagonal.

1.3 The Single Index Model and Portfolios

Suppose that the single index model (1) describes the returns on two assets. That is,

$$\begin{aligned} R_{1t} &= \alpha_1 + \beta_{1,M} R_{Mt} + \varepsilon_{1t}, && (2) \\ R_{2t} &= \alpha_2 + \beta_{2,M} R_{Mt} + \varepsilon_{2t}. && (3) \end{aligned}$$

Consider forming a portfolio of these two assets. Let x_1 denote the share of wealth in asset 1, x_2 the share of wealth in asset 2 and suppose that $x_1 + x_2 = 1$. The return on this portfolio using (2) and (3) is then

$$
\begin{aligned}
R_{pt} &= x_1 R_{1t} + x_2 R_{2t} \\
&= x_1(\alpha_1 + \beta_{1,M} R_{Mt} + \varepsilon_{1t}) + x_2(\alpha_2 + \beta_{2,M} R_{Mt} + \varepsilon_{2t}) \\
&= (x_1\alpha_1 + x_2\alpha_2) + (x_1\beta_{1,M} + x_2\beta_{2,M}) R_{Mt} + (x_1\varepsilon_{1t} + x_2\varepsilon_{2t}) \\
&= \alpha_p + \beta_{p,M} R_{Mt} + \varepsilon_{pt}
\end{aligned}
$$

where $\alpha_p = x_1\alpha_1 + x_2\alpha_2$, $\beta_{p,M} = x_1\beta_{1,M} + x_2\beta_{2,M}$ and $\varepsilon_{pt} = x_1\varepsilon_{1t} + x_2\varepsilon_{2t}$. Hence, the single index model will hold for the return on the portfolio where the parameters of the single index model are weighted averages of the parameters of the individual assets in the portfolio. In particular, the beta of the portfolio is a weighted average of the individual betas where the weights are the portfolio weights.

Example 2 *To be completed*

The additivity result of the single index model above holds for portfolios of any size. To illustrate, suppose the single index model holds for a collection of N assets:

$$R_{it} = \alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it} \ (i = 1, \ldots, N)$$

Consider forming a portfolio of these N assets. Let x_i denote the share of wealth invested in asset i and assume that $\sum_{i=1}^N = 1$. Then the return on the portfolio is

$$
\begin{aligned}
R_{pt} &= \sum_{i=1}^N x_i(\alpha_i + \beta_{i,M} R_{Mt} + \varepsilon_{it}) \\
&= \sum_{i=1}^N x_i\alpha_i + \left(\sum_{i=1}^N x_i\beta_{i,M}\right) R_{Mt} + \sum_{i=1}^N x_i\varepsilon_{it} \\
&= \alpha_p + \beta_p R_{Mt} + \varepsilon_{pt}
\end{aligned}
$$

where $\alpha_p = \sum_{i=1}^N x_i\alpha_i$, $\beta_p = \left(\sum_{i=1}^N x_i\beta_{i,M}\right)$ and $\varepsilon_{pt} = \sum_{i=1}^N x_i\varepsilon_{it}$.

1.3.1 The Single Index Model and Large Portfolios

To be completed

2 "Beta" as a Measure of portfolio Risk

A key insight of portfolio theory is that, due to diversification, the risk of an individual asset should be based on how it affects the risk of a well diversified portfolio if it is added to the portfolio. The preceding section illustrated that individual specific risk, as measured by the asset's own variance, can be diversified away in large well diversified portfolios whereas the covariances of the asset with the other assets in the portfolio cannot be completely diversi&ed away. The so-called "beta" of an asset captures this covariance contribution and so is a measure of the contribution of the asset to overall portfolio variability.

To illustrate, consider an equally weighted portfolio of 99 stocks and let R_{99} denote the return on this portfolio and σ_{99}^2 denote the variance. Now consider adding one stock, say IBM, to the portfolio. Let R_{IBM} and σ_{IBM}^2 denote the return and variance of IBM and let $\sigma_{99,IBM} = cov(R_{99}, R_{IBM})$. What is the contribution of IBM to the risk, as measured by portfolio variance, of the portfolio? Will the addition of IBM make the portfolio riskier (increase portfolio variance)? Less risky (decrease portfolio variance)? Or have no effect (not change portfolio variance)? To answer this question, consider a new equally weighted portfolio of 100 stocks constructed as

$$R_{100} = (0.99) \cdot R_{99} + (0.01) \cdot R_{IBM}.$$

The variance of this portfolio is

$$\begin{aligned}
\sigma_{100}^2 &= var(R_{100}) = (0.99)^2\sigma_{99}^2 + (0.01)^2\sigma_{IBM}^2 + 2(0.99)(0.01)\sigma_{99,IBM} \\
&= (0.98)\sigma_{99}^2 + (0.0001)\sigma_{IBM}^2 + (0.02)\sigma_{99,IBM} \\
&\approx (0.98)\sigma_{99}^2 + (0.02)\sigma_{99,IBM}.
\end{aligned}$$

Now if

- $\sigma_{100}^2 = \sigma_{99}^2$ then adding IBM does not change the variability of the portfolio;
- $\sigma_{100}^2 > \sigma_{99}^2$ then adding IBM increases the variability of the portfolio;
- $\sigma_{100}^2 < \sigma_{99}^2$ then adding IBM decreases the variability of the portfolio.

Consider the &rst case where $\sigma_{100}^2 = \sigma_{99}^2$. This implies (approximately) that

$$(0.98)\sigma_{99}^2 + (0.02)\sigma_{99,IBM} = \sigma_{99}^2$$

which upon rearranging gives the condition

$$\frac{\sigma_{99,IBM}}{\sigma_{99}^2} = \frac{cov(R_{99}, R_{IBM})}{var(R_{99})} = 1$$

De&ning

$$\beta_{99,IBM} = \frac{cov(R_{99}, R_{IBM})}{var(R_{99})}$$

then adding IBM does not change the variability of the portfolio as long as $\beta_{99,IBM} = 1$. Similarly, it is easy to see that $\sigma_{100}^2 > \sigma_{99}^2$ implies that $\beta_{99,IBM} > 1$ and $\sigma_{100}^2 < \sigma_{99}^2$ implies that $\beta_{99,IBM} < 1$.

In general, let R_p denote the return on a large diversi&ed portfolio and let R_i denote the return on some asset i. Then

$$\beta_{p,i} = \frac{cov(R_p, R_i)}{var(R_p)}$$

measures the contribution of asset i to the overall risk of the portfolio.

2.1 The single index model and Portfolio Theory

To be completed

2.2 Estimation of the single index model by Least Squares Regression

Consider a sample of size T of observations on R_{it} and R_{Mt}. We use the lower case variables r_{it} and r_{Mt} to denote these observed values. The method of least squares &nds the □best &tting□line to the scatter-plot of data as follows. For a given estimate of the best &tting line

$$\widehat{r}_{it} = \widehat{\alpha}_i + \widehat{\beta}_{i,M} r_{Mt}, \ t = 1, \ldots, T$$

create the T observed errors

$$\widehat{\varepsilon}_{it} = r_{it} - \widehat{r}_{it} = r_{it} - \widehat{\alpha}_i - \widehat{\beta}_{i,M} r_{Mt}, \ t = 1, \ldots, T$$

Now some lines will &t better for some observations and some lines will &t better for others. The least squares regression line is the one that minimizes the error sum of squares (ESS)

$$SSR(\widehat{\alpha}_i, \widehat{\beta}_{i,M}) = \sum_{t=1}^{T} \widehat{\varepsilon}_{it}^2 = \sum_{t=1}^{T} (r_{it} - \widehat{\alpha}_i - \widehat{\beta}_{i,M} r_{Mt})^2$$

The minimizing values of $\widehat{\alpha}_i$ and $\widehat{\beta}_{i,M}$ are called the (ordinary) least squares (OLS) estimates of α_i and $\beta_{i,M}$. Notice that $SSR(\widehat{\alpha}_i, \widehat{\beta}_{i,M})$ is a quadratic function in $(\widehat{\alpha}_i, \widehat{\beta}_{i,M})$ given the data and so the minimum values can be easily obtained using calculus. The &rst order conditions for a minimum are

$$\begin{aligned}
0 &= \frac{\partial SSR}{\partial \widehat{\alpha}_i} = -2 \sum_{t=1}^{T} (r_{it} - \widehat{\alpha}_i - \widehat{\beta}_{i,M} r_{Mt}) = -2 \sum_{t=1}^{T} \widehat{\varepsilon}_{it} \\
0 &= \frac{\partial SSR}{\partial \widehat{\beta}_{i,M}} = -2 \sum_{t=1}^{T} (r_{it} - \widehat{\alpha}_i - \widehat{\beta}_{i,M} r_{Mt}) r_{Mt} = -2 \sum_{t=1}^{T} \widehat{\varepsilon}_{it} r_{Mt}
\end{aligned}$$

which can be rearranged as

$$\begin{aligned}
\sum_{t=1}^{T} r_{it} &= T\widehat{\alpha}_i + \widehat{\beta}_{i,M} \sum_{t=1}^{T} r_{Mt} \\
\sum_{t=1}^{T} r_{it} r_{Mt} &= \widehat{\alpha}_i \sum_{t=1}^{T} r_{Mt} + \widehat{\beta}_{i,M} \sum_{t=1}^{T} r_{Mt}^2
\end{aligned}$$

These are two linear equations in two unknowns and by straightforward substitution the solution is

$$\begin{aligned}\widehat{\alpha}_i &= \bar{r}_i - \widehat{\beta}_{i,M}\bar{r}_M \\ \widehat{\beta}_{i,M} &= \frac{\sum_{t=1}^{T}(r_{it}-\bar{r}_i)(r_{Mt}-\bar{r}_M)}{\sum_{t=1}^{T}(r_{Mt}-\bar{r}_M)^2}\end{aligned}$$

where

$$\bar{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it},\ \ \bar{r}_M = \frac{1}{T}\sum_{t=1}^{T} r_{Mt}.$$

The equation for $\widehat{\beta}_{i,M}$ can be rewritten slightly to show that $\widehat{\beta}_{i,M}$ is a simple function of variances and covariances. Divide the numerator and denominator of the expression for $\widehat{\beta}_{i,M}$ by $\frac{1}{T-1}$ to give

$$\widehat{\beta}_{i,M} = \frac{\frac{1}{T-1}\sum_{t=1}^{T}(r_{it}-\bar{r}_i)(r_{Mt}-\bar{r}_M)}{\frac{1}{T-1}\sum_{t=1}^{T}(r_{Mt}-\bar{r}_M)^2} = \frac{\widehat{cov}(R_{it}, R_{Mt})}{\widehat{var}(R_{Mt})}$$

which shows that $\widehat{\beta}_{i,M}$ is the ratio of the estimated covariance between R_{it} and R_{Mt} to the estimated variance of R_{Mt}.

The least squares estimate of $\sigma^2_{\varepsilon,i} = var(\varepsilon_{it})$ is given by

$$\widehat{\sigma}^2_{\varepsilon,i} = \frac{1}{T-2}\sum_{t=1}^{T}\widehat{e}^2_{it} = \frac{1}{T-2}\sum_{t=1}^{T}(r_t - \widehat{\alpha}_i - \widehat{\beta}_{i,M} r_{Mt})^2$$

The divisor $T-2$ is used to make $\widehat{\sigma}^2_{\varepsilon,i}$ an unbiased estimator of $\sigma^2_{\varepsilon,\iota}$.

The least squares estimate of R^2 is given by

$$\widehat{R}^2_i = \frac{\widehat{\beta}^2_{i,M}\widehat{\sigma}^2_M}{\widehat{var}(R_{it})} = 1 - \frac{\widehat{\sigma}^2_{\varepsilon,i}}{\widehat{var}(R_{it})},$$

where

$$\widehat{var}(R_{it}) = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it}-\bar{r}_i)^2,$$

and gives a measure of the goodness of &t of the regression equation. Notice that $\widehat{R}^2_i = 1$ whenever $\widehat{\sigma}^2_{\varepsilon,i} = 0$ which occurs when $\widehat{\varepsilon}_{it} = 0$ for all values of t. In other words, $\widehat{R}^2_i = 1$ whenever the regression line has a perfect &t. Conversely, $\widehat{R}^2_i = 0$ when $\widehat{\sigma}^2_{\varepsilon,i} = \widehat{var}(R_{it})$; that is, when the market does not explain any of the variability of R_{it}. In this case, the regression has the worst possible &t.

3 Hypothesis Testing in the Single Index Model

3.1 A Review of Hypothesis Testing Concepts

To be completed.

3.2 Testing the Restriction $\alpha = 0$.

Using the single index model regression,

$$\begin{aligned} R_t &= \alpha + \beta R_{Mt} + \varepsilon_t,\ t = 1, ..., T \\ \varepsilon_t &\sim iid\ N(0, \sigma_\varepsilon^2),\ \varepsilon_t \text{ is independent of } R_{Mt} \end{aligned} \tag{4}$$

consider testing the null or maintained hypothesis $\alpha = 0$ against the alternative that $\alpha \neq 0$

$$H_0 : \alpha = 0\ vs.\ H_1 : \alpha \neq 0.$$

If H_0 is true then the single index model regression becomes

$$R_t = \beta R_{Mt} + \varepsilon_t$$

and $E[R_t|R_{Mt} = r_{Mt}] = \beta r_{Mt}$. We will reject the null hypothesis, $H_0 : \alpha = 0$, if the estimated value of α is either much larger than zero or much smaller than zero. Assuming $H_0 : \alpha = 0$ is true, $\hat{\alpha} \sim N(0, SE(\hat{\alpha})^2)$ and so is fairly unlikely that $\hat{\alpha}$ will be more than 2 values of $SE(\hat{\alpha})$ from zero. To determine how big the estimated value of α needs to be in order to reject the null hypothesis we use the t-statistic

$$t_{\alpha=0} = \frac{\widehat{\alpha} - 0}{\widehat{SE}(\widehat{\alpha})},$$

where $\widehat{\alpha}$ is the least squares estimate of α and $\widehat{SE}(\widehat{\alpha})$ is its estimated standard error. The value of the t-statistic, $t_{\alpha=0}$, gives the number of estimated standard errors that $\widehat{\alpha}$ is from zero. If the absolute value of $t_{\alpha=0}$ is much larger than 2 then the data cast considerable doubt on the null hypothesis $\alpha = 0$ whereas if it is less than 2 the data are in support of the null hypothesis[3]. To determine how big $|\ t_{\alpha=0}|$ needs to be to reject the null, we use the fact that under the statistical assumptions of the single index model and *assuming the null hypothesis is true*

$$t_{\alpha=0} \sim Student - t \text{ with } T - 2 \text{ degrees of freedom}$$

If we set the signi&cance level (the probability that we reject the null given that the null is true) of our test at, say, 5% then our decision rule is

$$\text{Reject } H_0 : \alpha = 0 \text{ at the 5\% level if } |t_{\alpha=0}| > |t_{T-2}(0.025)|$$

where t_{T-2} is the $2\frac{1}{2}\%$ critical value (quantile) from a Student-t distribution with $T - 2$ degrees of freedom.

Example 3 *single index model Regression for IBM*

[3] This interpretation of the t-statistic relies on the fact that, assuming the null hypothesis is true so that $\alpha = 0$, $\widehat{\alpha}$ is normally distributed with mean 0 and estimated variance $\widehat{SE}(\widehat{\alpha})^2$.

Consider the estimated MM regression equation for IBM using monthly data from January 1978 through December 1982:

$$\widehat{R}_{IBM,t} = \underset{(0.0068)}{-0.0002} + \underset{(0.0888)}{0.3390} \cdot R_{Mt},\ R^2 = 0.20,\ \widehat{\sigma}_{\varepsilon} = 0.0524$$

where the estimated standard errors are in parentheses. Here $\widehat{\alpha} = -0.0002$, which is very close to zero, and the estimated standard error, $\widehat{SE}(\hat{\alpha}) = 0.0068$, is much larger than $\widehat{\alpha}$. The t-statistic for testing $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$ is

$$t_{\alpha=0} = \frac{-0.0002 - 0}{0.0068} = -0.0363$$

so that $\widehat{\alpha}$ is only 0.0363 estimated standard errors from zero. Using a 5% signi&cance level, $|t_{58}(0.025)| \approx 2$ and

$$|t_{\alpha=0}| = 0.0363 < 2$$

so we do not reject $H_0 : \alpha = 0$ at the 5% level.

3.3 Testing Hypotheses about β

In the single index model regression β measures the contribution of an asset to the variability of the market index portfolio. One hypothesis of interest is to test if the asset has the same level of risk as the market index against the alternative that the risk is different from the market:

$$H_0 : \beta = 1 \ vs.\ H_1 : \beta \neq 1.$$

The data cast doubt on this hypothesis if the estimated value of β is much different from one. This hypothesis can be tested using the t-statistic

$$t_{\beta=1} = \frac{\widehat{\beta} - 1}{\widehat{SE}(\widehat{\beta})}$$

which measures how many estimated standard errors the least squares estimate of β is from one. The null hypothesis is reject at the 5% level, say, if $|t_{\beta=1}| > |t_{T-2}(0.025)|$. Notice that this is a two-sided test.

Alternatively, one might want to test the hypothesis that the risk of an asset is strictly less than the risk of the market index against the alternative that the risk is greater than or equal to that of the market:

$$H_0 : \beta = 1 \ vs.\ H_1 : \beta \geq 1.$$

Notice that this is a one-sided test. We will reject the null hypothesis only if the estimated value of β much greater than one. The t-statistic for testing this null hypothesis is the same as before but the decision rule is different. Now we reject the null at the 5% level if

$$t_{\beta=1} < -t_{T-2}(0.05)$$

where $t_{T-2}(0.05)$ is the one-sided 5% critical value of the Student-t distribution with $T-2$ degrees of freedom.

Example 4 *Single Index Regression for IBM cont'd*

Continuing with the previous example, consider testing $H_0 : \beta = 1$ *vs.* $H_1 : \beta \neq 1$. Notice that the estimated value of β is 0.3390, with an estimated standard error of 0.0888, and is fairly far from the hypothesized value $\beta = 1$. The t-statistic for testing $\beta = 1$ is

$$t_{\beta=1} = \frac{0.3390 - 1}{0.0888} = -7.444$$

which tells us that $\widehat{\beta}$ is more than 7 estimated standard errors below one. Since $t_{0.025,58} \approx 2$ we easily reject the hypothesis that $\beta = 1$.

Now consider testing $H_0 : \beta = 1$ *vs.* $H_1 : \beta \geq 1$. The t-statistic is still -7.444 but the critical value used for the test is now $-t_{58}(0.05) \approx -1.671$. Clearly, $t_{\beta=1} = -7.444 < -1.671 = -t_{58}(0.05)$ so we reject this hypothesis.

4 Estimation of the single index model: An Extended Example

Now we illustrate the estimation of the single index model using monthly data on returns over the ten year period January 1978 - December 1987. As our dependent variable we use the return on IBM and as our market index proxy we use the CRSP value weighted composite monthly return index based on transactions from the New York Stock Exchange and the American Stock Exchange. Let r_t denote the monthly return on IBM and r_{Mt} denote the monthly return on the CRSP value weighted index. Time plots of these data are given in figure 1 below.

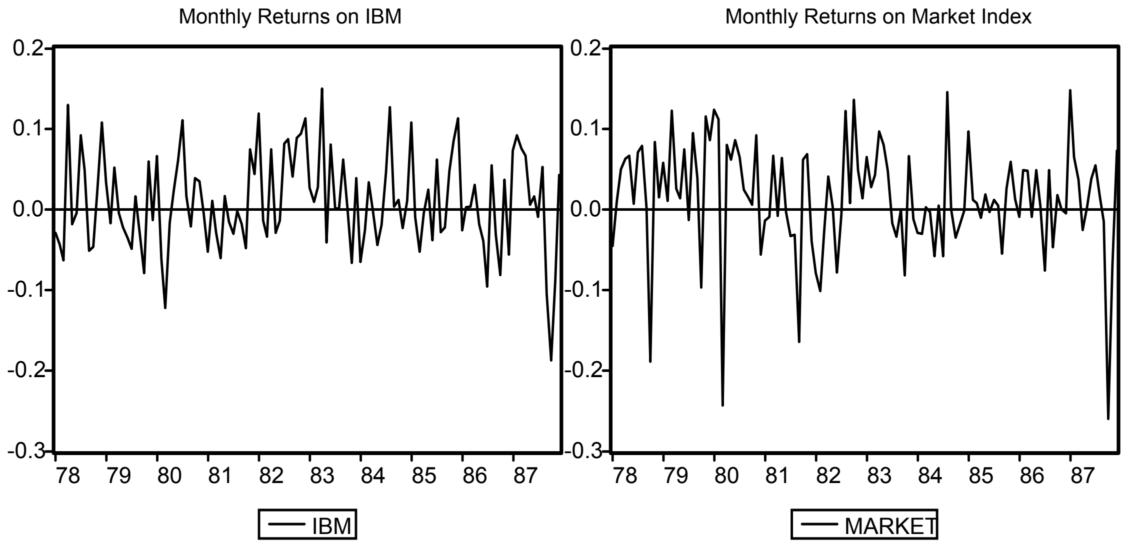


Figure 1

Notice that the IBM and the market index have similar behavior over the sample with the market index looking a little more volatile than IBM. Both returns dropped sharply during the October 1987 crash but there were a few times that the market dropped sharply whereas IBM did not. Sample descriptive statistics for the returns are displayed in &gure 2.

The mean monthly returns on IBM and the market index are 0.9617% and 1.3992% per month and the sample standard deviations are 5.9024% and 6.8353% per month, respectively.. Hence the market index on average had a higher monthly return and more volatility than IBM.

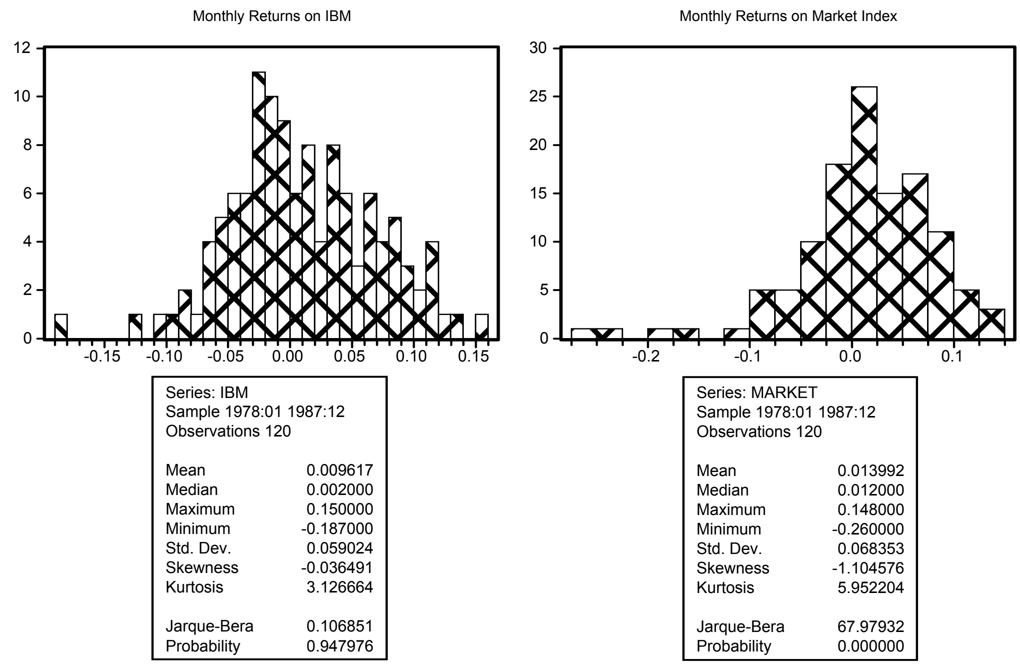


Figure 2

Notice that the histogram of returns on the market are heavily skewed left whereas the histogram for IBM is much more sysingle index modeletric about the mean. Also, the kurtosis for the market is much larger than 3 (the value for normally distributed returns) and the kurtosis for IBM is just slightly larger than 3. The negative skewness and large kurtosis of the market returns is caused by several large negative returns. The Jarque-Bera statistic for the market returns is 67.97, with a p-value 0.0000, and so we can easily reject the hypothesis that the market data are normally distributed. However, the Jarque-Bera statistic for IBM is only 0.1068, with a p-value of 0.9479, and we therefore cannot reject the hypothesis of normality.

The single index model regression is

$$R_t = \alpha + \beta R_{Mt} + \varepsilon_t,\ t = 1, \ldots, T$$

where it is assumed that $\varepsilon_t \sim iid\ N(0, \sigma^2)$ and is independent of R_{Mt}. We estimate this regression using the &rst &ve years of data from January 1978 - December 1982. In practice the single index model is seldom estimated using data covering more than &ve years because it is felt that β may change through time. The computer printout from Eviews is given in &gure 3 below

LS // Dependent Variable is IBM
Date: 02/12/98 Time: 11:41
Sample: 1978:01 1982:12
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.005312	0.006930	0.766533	0.4465
MARKET	0.327799	0.088987	3.683665	0.0005

R-squared	0.189598	Mean dependent var	0.011633
Adjusted R-squared	0.175625	S.D. dependent var	0.057283
S.E. of regression	0.052010	Akaike info criterion	-5.879874
Sum squared resid	0.156892	Schwarz criterion	-5.810063
Log likelihood	93.25991	F-statistic	13.56939
Durbin-Watson stat	1.577416	Prob(F-statistic)	0.000507

Figure 3

4.1 Explanation of Computer Output

The the items under the column labeled Variable are the variables in the estimated regression model. The variable "C" refers to the intercept in the regression and "MARKET" refers to r_{Mt}. The least squares regression coefficients are reported in the column labeled "Coefficient" and the estimated standard errors for the coefficients are in then next column. A standard way of reporting the estimated equation is

$$\widehat{r}_t = \underset{(0.0069)}{0.0053} + \underset{(0.0890)}{0.3278} \cdot r_{Mt}$$

where the estimated standard errors are reported underneath the estimated coefficients. The estimated intercept is close to zero at 0.0053, with a standard error of 0.0069 $(= \widehat{SE}(\widehat{\alpha}))$, and the estimated value of β is 0.3278, with an standard error of 0.0890 $(= \widehat{SE}(\widehat{\beta}))$. Notice that the estimated standard error of $\widehat{\beta}$ is much smaller than the estimated coefficient and indicates that β is estimated reasonably precisely. The estimated regression equation is displayed graphically in &gure 4 below.

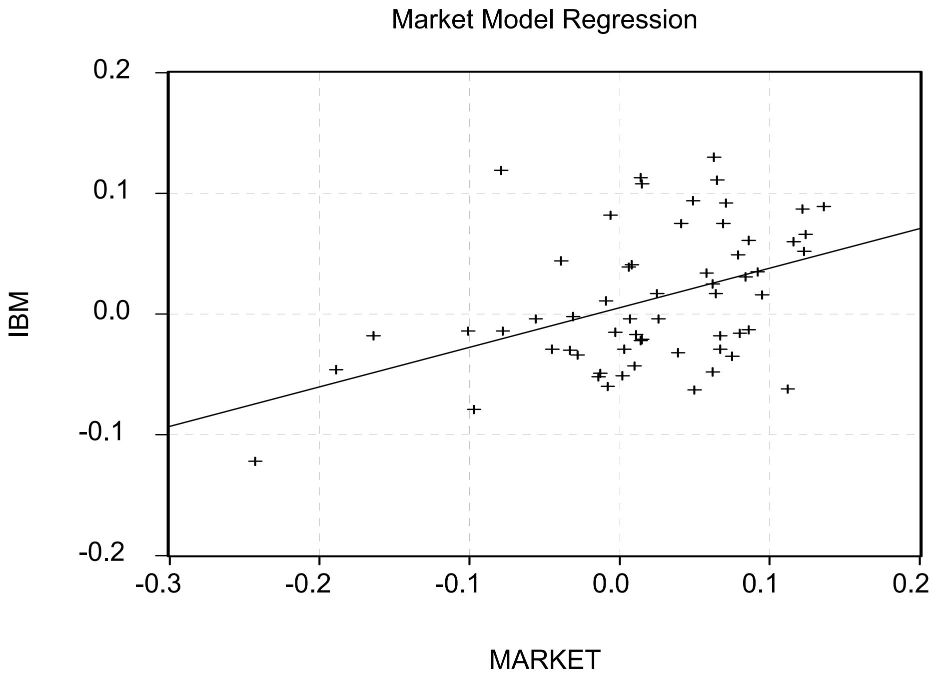


Figure 4

To evaluate the overall &t of the single index model regression we look at the R^2 of the regression, which measures the percentage of variability of R_t that is attributable to the variability in R_{Mt}, and the estimated standard deviation of the residuals, $\widehat{\sigma}_\varepsilon$. From the table, $R^2 = 0.190$ so the market index explains only 19% of the variability of IBM and 81% of the variability is not explained by the market. In the single index model regression, we can also interpret R^2 as the proportion of market risk in IBM and $1 - R^2$ as the proportion of &rm speci&c risk. The standard error (S.E.) of the regression is the square root of the least squares estimate of $\sigma_\varepsilon^2 = var(\varepsilon_t)$. From the above table, $\widehat{\sigma}_\varepsilon = 0.052$. Recall, ε_t captures the &rm speci&c risk of IBM and so $\widehat{\sigma}_\varepsilon$ is an estimate of the typical magnitude of the &rm speci&c risk. In order to interpret the magnitude of $\widehat{\sigma}_\varepsilon$ it is useful to compare it to the estimate of the standard deviation of R_t, which measures the total risk of IBM. This is reported in the table by the standard deviation (S.D.) of the dependent variable which equals 0.057. Notice that $\widehat{\sigma}_\varepsilon = 0.052$ is only slightly smaller than 0.057 so that the &rm speci&c risk is a large proportion of total risk (which is also reported by $1 - R^2$).

Con&dence intervals for the regression parameters are easily computed using the reported regression output. Since ε_t is assumed to be normally distributed 95% con&dence intervals for α and β take the form

$$\widehat{\alpha} \pm 2 \cdot \widehat{SE}(\widehat{\alpha})$$
$$\widehat{\beta} \pm 2 \cdot \widehat{SE}(\widehat{\beta})$$

The 95% confidence intervals are then

$$\begin{aligned}\alpha &: 0.0053 \pm 2 \cdot 0.0069 = [-.0085, 0.0191] \\ \beta &: 0.3278 \pm 2 \cdot 0.0890 = [0.1498, 0.5058]\end{aligned}$$

Our best guess of α is 0.0053 but we wouldn't be too surprised if it was as low as -0.0085 or as high as 0.0191. Notice that there are both positive and negative values in the confidence interval. Similarly, our best guess of β is 0.3278 but it could be as low as 0.1498 or as high as 0.5058. This is a fairly wide range given the interpretation of β as a risk measure. The interpretation of these intervals are as follows. In repeated samples, 95% of the time the estimated confidence intervals will cover the true parameter values.

The t-statistic given in the computer output is calculated as

$$t\text{-statistic} = \frac{\text{estimated coefficient} - 0}{\text{std. error}}$$

and it measures how many estimated standard errors the estimated coefficient is away from zero. This t-statistic is often referred to as a *basic significance* test because it tests the null hypothesis that the value of the true coefficient is zero. If an estimate is several standard errors from zero, so that it's t-statistic is greater than 2, then it is a good bet that the true coefficient is not equal to zero. From the data in the table, the t-statistic for α is 0.767 so that $\widehat{\alpha} = 0.0053$ is 0.767 standard errors from zero. Hence it is quite likely that the true value of α equals zero. The t-statistic for β is 3.684, $\widehat{\beta}$ is more than 3 standard errors from zero, and so it is very unlikely that $\beta = 0$. The Prob Value (p-value of the t-statistic) in the table gives the likelihood (computed from the Student-t curve) that, *given the true value of the coefficient is zero*, the data would generate the observed value of the t-statistic. The p-value for the t-statistic testing $\alpha = 0$ is 0.4465 so that it is quite likely that $\alpha = 0$. Alternatively, the p-value for the t-statistic testing $\beta = 0$ is 0.001 so it is very unlikely that $\beta = 0$.

4.2 Analysis of the Residuals

The single index model regression makes the assumption that $\varepsilon_t \sim iid\ N(0, \sigma_\varepsilon^2)$. That is the errors are independent and identically distributed with mean zero, constant variance σ_ε^2 and are normally distributed. It is always a good idea to check the behavior of the estimated residuals, $\widehat{\varepsilon}_t$, and see if they share the assumed properties of the true residuals ε_t. The figure below plots r_t (the actual data), $\widehat{r}_t = \widehat{\alpha} + \widehat{\beta} r_{Mt}$ (the fitted data) and $\widehat{\varepsilon}_t = r_t - \widehat{r}_t$ (the estimated residual data).

Market Model Regression for IBM

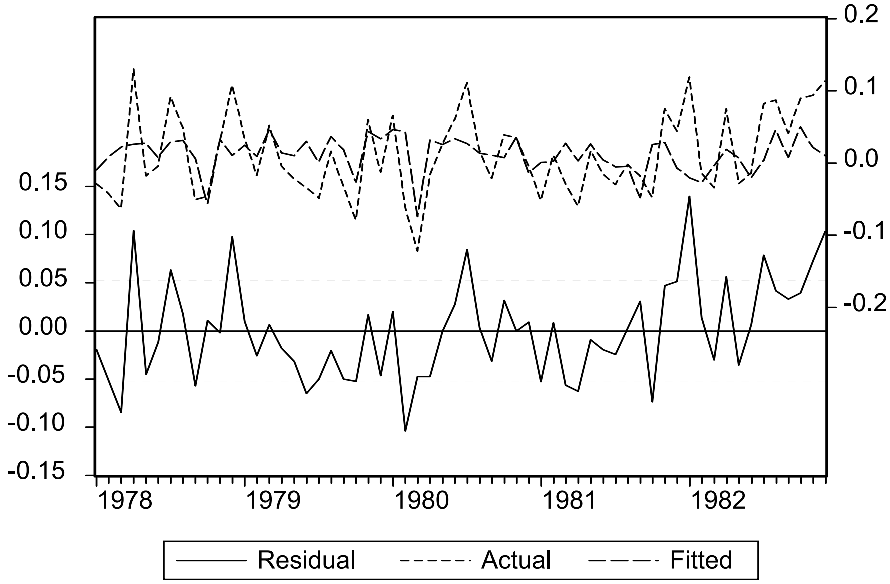


Figure 5

Notice that the &tted values do not track the actual values very closely and that the residuals are fairly large. This is due to low R^2 of the regression. The residuals appear to be fairly random by sight. We will develop explicit tests for randomness later on. The histogram of the residuals, displayed below, can be used to investigate the normality assumption. As a result of the least squares algorithm the residuals have mean zero as long as a constant is included in the regression. The standard deviation of the residuals is essentially equal to the standard error of the regression - the difference being due to the fact that the formula for the standard error of the regression uses $T - 2$ as a divisor for the error sum of squares and the standard deviation of the residuals uses the divisor $T - 1$.

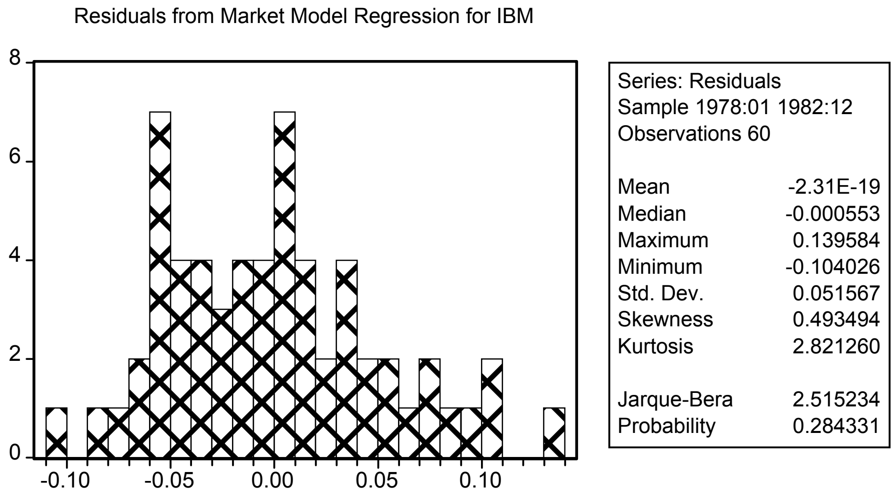


Figure 6

The skewness of the residuals is slightly positive and the kurtosis is a little less than 3. The hypothesis that the residuals are normally distributed can be tested using the Jarque-Bera statistic. This statistic is a function of the estimated skewness and kurtosis and is give by

$$JB = \frac{T}{6}\left(\widehat{S}^2 + \frac{(\widehat{K}-3)^2}{4}\right)$$

where $\widehat{S}$ denotes the estimated skewness and $\widehat{K}$ denotes the estimated kurtosis. If the residuals are normally distribued then $\widehat{S} \approx 0$ and $\widehat{K} \approx 3$ and $JB \approx 0$. Therefore, if $\widehat{S}$ is moderately different from zero or $\widehat{K}$ is much different from 3 then JB will get large and suggest that the data are not normally distributed. To determine how large JB needs to be to be able to reject the normality assumption we use the result that under the maintained hypothesis that the residuals are normally distributed JB has a chi-square distribution with 2 degrees of freedom:

$$JB \sim \chi^2_2.$$

For a test with signi&cance level 5%, the 5% right tail critical value of the chi-square distribution with 2 degrees of freedom, $\chi^2_2(0.05)$, is 5.99 so we would reject the null that the residuals are normally distributed if $JB > 5.99$. The Probability (p-value) reported by Eviews is the probability that a chi-square random variable with 2 degrees of freedom is greater than the observed value of JB :

$$P(\chi^2_2 \geq JB) = 0.2843.$$

For the IBM residuals this p-value is reasonably large and so there is not much data evidence against the normality assumption. If the p-value was very small, e.g., 0.05 or smaller, then the data would suggest that the residuals are not normally distributed.

1. The Capital Asset Pricing Model

The capital asset pricing model (CAPM) is an equilibrium model for expected returns and relies on a set of rather strict assumptions.

CAPM Assumptions

1. Many investors who are all price takers
2. All investors plan to invest over the same time horizon
3. There are no taxes or transactions costs
4. Investors can borrow and lend at the same risk-free rate over the planned investment horizon
5. Investors only care about expected return and variance. Investors like expected return but dislike variance. (A sufficient condition for this is that returns are all normally distributed)
6. All investors have the same information and beliefs about the distribution of returns
7. The market portfolio consists of all publicly traded assets

The implications of these assumptions are as follows

1. All investors use the Markowitz algorithm to determine the same set of efficient portfolios. That is, the efficient portfolios are combinations of the risk-free asset and the tangency portfolio and everyone's determination of the tangency portfolio is the same.
2. Risk averse investors put a majority of wealth in the risk-free asset (i.e. lend at the risk-free rate) whereas risk tolerant investors borrow at the risk-free rate and leverage their holdings of the tangency portfolio. In equilibrium total borrowing and lending must equalize so that the risk-free asset is in zero net supply when we aggregate across all investors.
3. Since everyone holds the same tangency portfolio and the risk-free asset is in zero net supply in the aggregate, when we aggregate over all investors the aggregate demand for assets is simply the tangency portfolio. The supply of all assets is simply the market portfolio (where the weight of an asset

in the market portfolio is just the market value of the asset divided by the total market value of all assets) and in equilibrium supply equal demand. *Therefore, in equilibrium the tangency portfolio is the market portfolio.*

4. Since the market portfolio is the tangency portfolio and the tangency portfolio is (mean-variance) efficient the market portfolio is also (mean-variance) efficient.

5. Since the market portfolio is efficient and there is a risk-free asset the security market line (SML) pricing relationship holds for all assets (and portfolios)

$$E[R_i] = r_r + \beta(E[R_{Mt}] - r_f)$$

or

$$\mu_i = r_f + \beta(\mu_M - r_f)$$

where R_i denotes the return on any asset or portfolio i, R_M denotes the return on the market portfolio and $\beta = cov(R_i, R_M)/var(R_m)$. The SML says that there is a linear relationship between the expected return on an asset and the "beta" of that asset with the market portfolio. Given a value for the market risk premium, $E[R] - r_r > 0$, the higher the beta on an asset the higher the expected return on the asset and vice-versa.

The SML relationship can be rewritten in terms of risk premia by simply subtracting r_f from both side of the SML equation:

$$E[R_i] - r_r = \beta(E[R_{Mt}] - r_f)$$

or

$$\mu_i - r_f = \beta(\mu_M - r_f)$$

and this linear relationship is illustrated graphically in figure 1. In terms of risk premia, the SML intersects the vertical axis at zero and has slope equal to $\mu_M - r_f$, the risk premium on the market portfolio (which is assumed to be positive). Low beta assets (less than 1) have risk premia less than the market and high beta (greater than 1) assets have risk premia greater than the market.

1.1. A Simple Regression Test of the CAPM

The SML relationship allows a test of the CAPM using a modified version of the market model regression equation. To see this, consider the *excess return market model* regression equation

$$\begin{aligned} R_t - r_f &= \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t, \ t = 1, ..., T \\ \varepsilon_t &\sim iid\ N(0, \sigma^2), \ \varepsilon_t \text{ is independent of } R_{Mt} \end{aligned} \quad (1.1)$$

where R_t denotes the return on any asset or portfolio and R_{Mt} is the return on some proxy for the market portfolio. Taking expectations of both sides of the excess return market model regression gives

$$E[R_t] - r_f = \alpha + \beta(E[R_{Mt}] - r_f)$$

and from the SML we see that the CAPM imposes the restriction

$$\alpha = 0$$

for every asset or portfolio. A simple testing strategy is as follows

- Estimate the excess return market model for every asset trades
- Test that $\alpha = 0$ in every regression

1.2. A Simple Prediction Test of the CAPM

Consider again the SML equation for the CAPM. The SML implies that there is a simple positive linear relationship between expected returns on any asset and the beta of that asset with the market portfolio. High beta assets have high expected returns and low beta assets have low expected returns. This linear relationship can be tested in the following way. Suppose we have a time series of returns on N assets (say 10 years of monthly data).

- Split a sample of time series data on returns into two equal sized subsamples.
- Estimate β for each asset in the sample using the first subsample of data. This gives N estimates of β.
- Using the second subsample of data, compute the average returns on the N assets (this is an estimate of $E[R_i] = \mu_i$). This give N estimates of μ.
- Plot the SML using the estimated betas and average returns and see if it intersects at zero on the vertical axis and has slope equal to the average risk premium on the market portfolio.

2. Hypothesis Testing using the Excess Return Market Model

In this section, we illustrate how to carry out some simple hypothesis tests concerning the parameters of the excess returns market model regression. Before we begin, we review some basic concepts from the theory of hypothesis testing.

2.1. Testing the CAPM Restriction $\alpha = 0$.

Using the market model regression,

$$\begin{aligned} R_t - r_f &= \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, ..., T \\ \varepsilon_t &\sim iid\ N(0, \sigma^2),\ \varepsilon_t \text{ is independent of } R_{Mt} \end{aligned} \tag{2.1}$$

consider testing the null or maintained hypothesis that the CAPM holds for an asset against the alternative hypothesis that the CAPM does not hold. These hypotheses can be formulated as the two-sided test

$$H_0 : \alpha = 0\ vs.\ H_1 : \alpha \neq 0.$$

We will reject the null hypothesis, $H_0 : \alpha = 0$, if the estimated value of α is either much larger than zero or much smaller than zero. To determine how big the estimated value of α needs to be in order to reject the CAPM we use the t-statistic

$$t_{\alpha=0} = \frac{\widehat{\alpha} - 0}{\widehat{SE}(\widehat{\alpha})},$$

where $\widehat{\alpha}$ is the least squares estimate of α and $\widehat{SE}(\widehat{\alpha})$ is its estimated standard error. The value of the t-statistic, $t_{\alpha=0}$, gives the number of estimated standard errors that $\widehat{\alpha}$ is from zero. If the absolute value of $t_{\alpha=0}$ is much larger than 2 then the data cast considerable doubt on the null hypothesis $\alpha = 0$ whereas if it is less than 2 the data are in support of the null hypothesis[1]. To determine how big $|\ t_{\alpha=0}|$ needs to be to reject the null, we use the fact that under the statistical assumptions of the market model and *assuming the null hypothesis is true*

$$t_{\alpha=0} \sim Student - t \text{ with } T - 2 \text{ degrees of freedom}$$

If we set the significance level (the probability that we reject the null given that the null is true) of our test at, say, 5% then our decision rule is

$$\text{Reject } H_0 : \alpha = 0 \text{ at the 5\% level if } |t_{\alpha=0}| > t_{0.025,T-2}$$

[1] This interpretation of the t-statistic relies on the fact that, assuming the null hypothesis is true so that $\alpha = 0$, $\widehat{\alpha}$ is normally distributed with mean 0 and estimated variance $\widehat{SE}(\widehat{\alpha})^2$.

where $t_{0.025,T-2}$ is the $2\frac{1}{2}\%$ critical value from a Student-t distribution with $T-2$ degrees of freedom.

Example 2.1. *CAPM Regression for IBM*

To illustrate the testing of the CAPM using the excess returns market model regression consider the regression output in figure 2

LS // Dependent Variable is IBMRP
Date: 02/23/98 Time: 15:50
Sample: 1978:01 1982:12
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000248	0.006836	-0.036301	0.9712
MARKETRP	0.339012	0.088799	3.817747	0.0003

R-squared	0.200829	Mean dependent var	0.003548
Adjusted R-squared	0.187050	S.D. dependent var	0.058103
S.E. of regression	0.052388	Akaike info criterion	-5.865397
Sum squared resid	0.159180	Schwarz criterion	-5.795586
Log likelihood	92.82561	F-statistic	14.57520
Durbin-Watson stat	1.566378	Prob(F-statistic)	0.000330

The estimated regression equation using monthly data from January 1978 through December 1982 is

$$\widehat{R_{IBM,t} - r_f} = \underset{(0.0068)}{-0.0002} + \underset{(0.0888)}{0.3390} \cdot (R_{M,t} - r_f),\ R^2 = 0.20,\ \hat{\sigma} = 0.0524$$

where the estimated standard errors are in parentheses. Here $\hat{\alpha} = -0.0002$, which is very close to zero, and the estimated standard error is 0.0068 is much larger than $\hat{\alpha}$. The t-statistic for testing $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$ is

$$t_{\alpha=0} = \frac{-0.0002 - 0}{0.0068} = -0.0363$$

so that $\widehat{\alpha}$ is only 0.0363 estimated standard errors from zero. Using a 5% significance level, $t_{0.025,58} \approx 2$ and

$$|t_{\alpha=0}| = 0.0363 < 2$$

so we do not reject $H_0 : \alpha = 0$ at the 5% level. Therefore, the CAPM appears to hold for IBM.

2.2. Testing Hypotheses about β

In the excess returns market model regression β measures the contribution of an asset to the variability of the market index portfolio. One hypothesis of interest is to test if the asset has the same level of risk as the market index against the alternative that the risk is different from the market:

$$H_0 : \beta = 1 \; vs. \; H_1 : \beta \neq 1.$$

The data cast doubt on this hypothesis if the estimated value of β is much different from one. This hypothesis can be tested using the t-statistic

$$t_{\beta=1} = \frac{\widehat{\beta} - 1}{\widehat{SE}(\widehat{\beta})}$$

which measures how many estimated standard errors the least squares estimate of β is from one. The null hypothesis is reject at the 5% level, say, if $|t_{\beta=1}| > t_{0.025,T-2}$. Notice that this is a two-sided test.

Alternatively, one might want to test the hypothesis that the risk of an asset is strictly less than the risk of the market index against the alternative that the risk is greater than or equal to that of the market:

$$H_0 : \beta < 1 \; vs. \; H_1 : \beta \geq 1.$$

Notice that this is a one-sided test. We will reject the null hypothesis only if the estimated value of β much greater than one. The t-statistic for testing this null hypothesis is the same as before but the decision rule is different. Now we reject the null at the 5% level if

$$t_{\beta=1} < -t_{0.05,T-2}$$

where $t_{0.05,T-2}$ is the one-sided 5% critical value of the Student-t distribution with $T - 2$ degrees of freedom.

Example 2.2. *CAPM Regression for IBM cont'd*

Continuing with the previous example, consider testing $H_0 : \beta = 1$ *vs.* $H_1 : \beta \neq 1$. Notice that the estimated value of β is 0.3390, with an estimated standard error of 0.0888, and is fairly far from the hypothesized value $\beta = 1$. The t-statistic for testing $\beta = 1$ is

$$t_{\beta=1} = \frac{0.3390 - 1}{0.0888} = -7.444$$

which tells us that $\widehat{\beta}$ is more than 7 estimated standard errors below one. Since $t_{0.025,58} \approx 2$ we easily reject the hypothesis that $\beta = 1$.

Now consider testing $H_0 : \beta < 1$ *vs.* $H_1 : \beta \geq 1$. The t-statistic is still -7.444 but the critical value used for the test is now $-t_{0.05,58} \approx -1.671$. Clearly, $t_{\beta=1} = -7.444 < -1.671 = -t_{0.05,58}$ so we reject this hypothesis.

2.3. Testing Joint Hypotheses about α and β

Often it is of interest to formulate hypothesis tests that involve both α and β. For example, consider the joint hypothesis that the CAPM holds and that an asset has the same risk as the market. The null hypothesis in this case can be formulated as

$$H_0 : \alpha = 0 \text{ and } \beta = 1.$$

The null will be rejected if either the CAPM doesn't hold, the asset has risk different from the market index or both. Thus the alternative is formulated as

$$H_1 : \alpha \neq 0, \text{or } \beta \neq 1 \text{ or } \alpha \neq 0 \text{ and } \beta \neq 1.$$

This type of joint hypothesis is easily tested using a so-called F-test. The idea behind the F-test is to estimate the model imposing the restrictions specified under the null hypothesis and compare the fit of the restricted model to the fit of the model with no restrictions imposed.

The fit of the unrestricted (UR) excess return market model is measured by the (unrestricted) error sum of squares (ESS)

$$ESS_{UR} = \sum_{t=1}^{T} \widehat{\varepsilon}_t^2 = \sum_{t=1}^{T} (R_t - r_f - \widehat{\alpha} - \widehat{\beta}(R_{Mt} - r_f))^2.$$

Recall, this is the quantity that is minimized during the least squares algorithm. Now, the excess return market model imposing the restrictions under H_0 is

$$\begin{aligned} R_t - r_f &= 0 + 1 \cdot (R_{Mt} - r_f) + \varepsilon_t \\ &= R_{Mt} - r_f + \varepsilon_t. \end{aligned}$$

Notice that there are no parameters to be estimated in this model which can be seen by subtracting $R_{Mt} - r_f$ from both sides of the restricted model to give

$$R_t - R_{Mt} = \widetilde{\varepsilon}_t$$

The fit of the restricted (R) model is then measured by the restricted error sum of squares

$$ESS_R = \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2 = \sum_{t=1}^{T} (R_t - R_{Mt})^2.$$

Now since the least squares algorithm works to minimize ESS, the restricted error sum of squares, ESS_R, must be at least as big as the unrestricted error sum of squares, ESS_{UR}. If the restrictions imposed under the null are true then $ESS_R \approx ESS_{UR}$ (with ESS_R always slightly bigger than ESS_{UR}) but if the restrictions are not true then ESS_R will be quite a bit bigger than ESS_{UR}. The F-statistic measures the (adjusted) percentage difference in fit between the restricted and unrestricted models and is given by

$$F = \frac{(ESS_R - ESS_{UR})/q}{ESS_{UR}/(T-k)} = \frac{(ESS_R - ESS_{UR})}{q \cdot \widehat{\sigma}_{UR}^2},$$

where q equals the number of restrictions imposed under the null hypothesis, k denotes the number of regression coefficients estimated under the unrestricted model and $\widehat{\sigma}_{UR}^2$ denotes the estimated variance of ε_t under the unrestricted model. Under the assumption that the null hypothesis is true, the F-statistic is distributed as an F random variable with q and $T - 2$ degrees of freedom:

$$F \sim F(q, T-2).$$

Notice that an F random variable is always positive since $ESS_R > ESS_{UR}$. The null hypothesis is rejected, say at the 5% significance level, if

$$F > F_{0.95}(q, T-2)$$

where $F_{0.95}(q, T-2)$ is the 95% quantile of the distribution of $F(q, T-2)$. For the hypothesis $H_0 : \alpha = 0$ and $\beta = 1$ there are $q = 2$ restrictions under the null and $k = 2$ regression coefficients estimated under the unrestricted model. The F-statistic is then

$$F_{\alpha=0,\beta=1} = \frac{(ESS_R - ESS_{UR})/2}{ESS_{UR}/(T-2)}$$

Example 2.3. *CAPM Regression for IBM cont'd*

Consider testing the hypothesis $H_0 : \alpha = 0$ and $\beta = 1$ for the IBM data. The unrestricted error sum of squares, ESS_{UR}, is obtained from the unrestricted regression output in figure 2 and is called **Sum Square Resid:**

$$ESS_{UR} = 0.159180.$$

To form the restricted sum of squared residuals, we create the new variable $\tilde{\varepsilon}_t = R_t - R_{Mt}$ and form the sum of squares $ESS_R = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2 = 0.31476$. Notice that $ESS_R > ESS_{UR}$. The F-statistic is then

$$F_{\alpha=0,\beta=1} = \frac{(0.31476 - 0.159180)/2}{0.159180/58} = 28.34.$$

The 95% quantile of the F-distribution with 2 and 58 degrees of freedom is about 3.15. Since $F_{\alpha=0,\beta=1} = 28.34 > 3.15 = F_{0.95}(2, 58)$ we reject $H_0 : \alpha = 0$ and $\beta = 1$ at the 5% level.

2.4. Testing the Stability of α and β over time

In many applications of the CAPM, β is estimated using past data and the estimated value of β is assumed to hold over some future time period. Since the characteristics of assets change over time it is of interest to know if β changes over time. To illustrate, suppose we have a ten year sample of monthly data ($T = 120$) on returns that we split into two five year subsamples. Denote the first five years as $t = 1, ..., T_B$ and the second five years as $t = T_{B+1}, ..., T$. The date $t = T_B$ is the "break date" of the sample and it is chosen arbitrarily in this context. Since the samples are of equal size (although they do not have to be) $T - T_B = T_B$. The excess returns market model regression which assumes that both α and β are constant over the entire sample is

$$\begin{aligned} R_t - r_f &= \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t, \ t = 1, \ldots, T \\ \varepsilon_t &\sim iid\ N(0, \sigma^2) \text{ independent of } R_{Mt}. \end{aligned}$$

There are two cases of interest: (1) β may differ over the two subsamples; (2) α and β may differ over the two subsamples.

2.4.1. Testing Structural Change in β only

If α is the same but β is different over the subsamples then we really have two excess return market model regressions

$$\begin{aligned} R_t - r_f &= \alpha + \beta_1(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, \ldots, T_B \\ R_t - r_f &= \alpha + \beta_2(R_{Mt} - r_f) + \varepsilon_t,\ t = T_{B+1}, \ldots, T \end{aligned}$$

that share the same intercept α but have different slopes $\beta_1 \neq \beta_2$. We can capture such a model very easily using a "step dummy variable" defined as

$$\begin{aligned} D_t &= 0,\ t \leq T_B \\ &= 1,\ t > T_B \end{aligned}$$

and re-writing the regression model as

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + D_t(R_{Mt} - r_f) + \varepsilon_t$$

The model for the first subsample when $D_t = 0$ is

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, \ldots, T_B$$

and the model for the second subsample when $D_t = 1$ is

$$\begin{aligned} R_t - r_f &= \alpha + \beta(R_{Mt} - r_f) + \delta(R_{Mt} - r_f) + \varepsilon_t,\ t = T_{B+1}, \ldots, T \\ &= \alpha + (\beta + \delta)(R_{Mt} - r_f) + \varepsilon_t. \end{aligned}$$

Notice that the "beta" in the first sample is $\beta_1 = \beta$ and the beta in the second subsample is $\beta_2 = \beta + \delta$. If $\delta < 0$ the second sample beta is smaller than the first sample beta and if $\delta > 0$ the beta is larger.

We can test the constancy of beta over time by testing whether $\delta = 0$:

H_0 : (beta is constant over time) $\delta = 0$ vs. H_1 : (beta is not constant over time) $\delta \neq 0$

The test statistic is simply the t-statistic

$$t_{\delta=0} = \frac{\widehat{\delta} - 0}{\widehat{SE}(\widehat{\delta})} = \frac{\widehat{\delta}}{\widehat{SE}(\widehat{\delta})}$$

and we reject the hypothesis $\delta = 0$ at the 5% level, say, if $|t_{\delta=0}| > t_{0.025,T-3}$.

Example 2.4. *CAPM Regression for IBM cont'd*

The Eviews output for the excess returns market model regression augmented with the structural change dummy is give in figure 3.

LS // Dependent Variable is IBMRP
Date: 02/25/98 Time: 16:37
Sample: 1978:01 1987:12
Included observations: 120

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.000136	0.004559	-0.029853	0.9762
MARKETRP	0.338801	0.083702	4.047689	0.0001
DUM*MARKETRP	0.315783	0.136575	2.312150	0.0225

R-squared	0.310824	Mean dependent var	0.002778
Adjusted R-squared	0.299043	S.D. dependent var	0.059299
S.E. of regression	0.049647	Akaike info criterion	-5.980965
Sum squared resid	0.288381	Schwarz criterion	-5.911278
Log likelihood	191.5853	F-statistic	26.38392
Durbin-Watson stat	1.943555	Prob(F-statistic)	0.000000

and the estimated equation is given by

$$\begin{aligned} \widehat{R_{IBM,t}} - r_f &= \underset{(0.0045)}{-0.0001} + \underset{(0.0837)}{0.3388} \cdot (R_{M,t} - r_f) + \underset{(0.1366)}{0.3158} \cdot D \cdot (R_{M,t} - r_f), \\ R^2 &= 0.311,\ \widehat{\sigma} = 0.0496. \end{aligned}$$

The estimated value of β is 0.3388, with a standard error of 0.0837, and the estimated value of δ is 0.3158, with a standard error of 0.1366. The t-statistic for testing $\delta = 0$ is given by

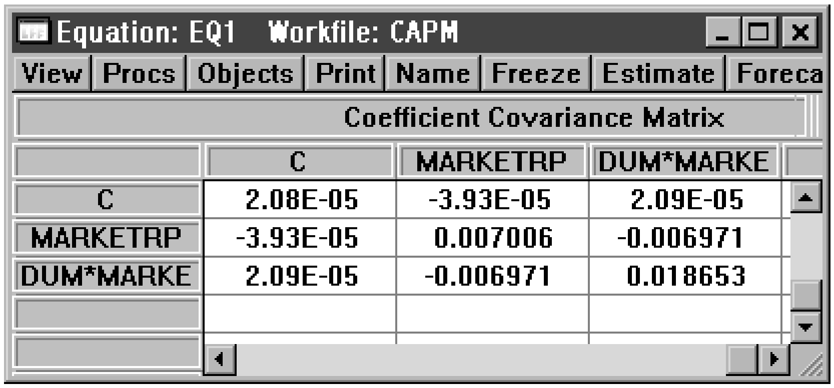
$$t_{\delta=0} = \frac{0.3158}{0.1366} = 2.312$$

which is greater than $t_{0.025,117} = 1.98$ so we reject the null hypothesis (at the 5% significance level) that beta is the same over the two subsamples.

The estimated value of beta over the second subsample is $\widehat{\beta} + \widehat{\delta} = 0.3388 + 0.3158 = 0.6546$. To get the estimated standard error for this estimate we note that

$$\widehat{var}(\widehat{\beta} + \widehat{\delta}) = \widehat{var}(\widehat{\beta}) + \widehat{var}(\widehat{\delta}) + 2 \cdot \widehat{cov}(\widehat{\beta}, \widehat{\delta})$$

and these numbers can be obtained from the elements of $\widehat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ where $\mathbf{X}$ is a $T \times 3$ matrix with elements $(1, R_{Mt} - r_f, D_t \cdot (R_{Mt} - r_f))$. Eviews computes this covariance matrix and it is displayed in figure 4.



	C	MARKETRP	DUM*MARKE
C	2.08E-05	-3.93E-05	2.09E-05
MARKETRP	-3.93E-05	0.007006	-0.006971
DUM*MARKE	2.09E-05	-0.006971	0.018653

From figure 4 we see that $\widehat{var}(\widehat{\beta}) = 0.007006$, $\widehat{var}(\widehat{\delta}) = 0.018653$ and $\widehat{cov}(\widehat{\beta}, \widehat{\delta}) = -0.006971$ so that

$$\widehat{var}(\widehat{\beta} + \widehat{\delta}) = 0.007006 + 0.018653 + 2 \cdot (-0.006971) = 0.011717$$

and

$$\widehat{SE}(\widehat{\beta} + \widehat{\delta}) = \sqrt{0.011717} = 0.1082.$$

2.4.2. Testing Structural Change in α and β

Now consider the case where both α and β are allowed to be different over the two subsamples:

$$\begin{aligned} R_t - r_f &= \alpha_1 + \beta_1(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, \ldots, T_B \\ R_t - r_f &= \alpha_2 + \beta_2(R_{Mt} - r_f) + \varepsilon_t,\ t = T_{B+1}, \ldots, T \end{aligned}$$

The dummy variable specification in this case is

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \delta_1 \cdot D_t + \delta_2 \cdot D_t(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, \ldots, T.$$

When $D_t = 0$ the model becomes

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t,\ t = 1, \ldots, T_B,$$

so that $\alpha_1 = \alpha$ and $\beta_1 = \beta$, and when $D_t = 1$ the model is

$$R_t - r_f = (\alpha + \delta_1) + (\beta + \delta_2)(R_{Mt} - r_f) + \varepsilon_t,\ t = T_{B+1}, \ldots, T,$$

so that $\alpha_2 = \alpha + \delta_1$ and $\beta_2 = \beta + \delta_2$. The hypothesis of no structural change is now

$$H_0 : \delta_1 = 0 \text{ and } \delta_2 = 0 \text{ vs. } H_1 : \delta_1 \neq 0 \text{ or } \delta_2 \neq 0 \text{ or } \delta_1 \neq 0 \text{ and } \delta_2 \neq 0.$$

The test statistic for this joint hypothesis is the F-statistic

$$F_{\delta_1=0,\delta_2=0} = \frac{(ESS_R - ESS_{UR})/2}{ESS_{UR}/(T-4)}$$

since there are two restrictions and four regression parameters estimated under the unrestricted model. The unrestricted (UR) model is the dummy variable regression that allows the intercepts and slopes to differ in the two subsamples and the restricted model (R) is the regression where these parameters are constrained to be the same in the two subsamples.

The unrestricted error sum of squares, ESS_{UR}, can be computed in two ways. The first way is based on the dummy variable regression. The second is based on estimating separate regression equations for the two subsamples and adding together the resulting error sum of squares. Let ESS_1 and ESS_2 denote the error sum of squares from separate regressions. Then

$$ESS_{UR} = ESS_1 + ESS_2.$$

Example 2.5. *CAPM regression for IBM cont'd*

The unrestricted regression (Eviews output not shown) is

$$\begin{aligned}
\widehat{R_{IBM,t} - r_f} &= \underset{(0.0065)}{-0.0001} + \underset{(0.0845)}{0.3388} \cdot (R_{Mt} - r_f) \\
&\quad + \underset{(0.0092)}{0.0002} \cdot D_t + \underset{(0.1377)}{0.3158} \cdot D_t(R_{Mt} - r_f), \\
R^2 &= 0.311,\ \widehat{\sigma} = 0.050,\ ESS_{UR} = 0.288379,
\end{aligned}$$

and the restricted regression is

$$\begin{aligned}
\widehat{R_{IBM,t} - r_f} &= \underset{(0.0046)}{-0.0005} + \underset{(0.0675)}{0.4568} \cdot (R_{Mt} - r_f), \\
R^2 &= 0.279,\ \widehat{\sigma} = 0.051,\ ESS_R = 0.301558.
\end{aligned}$$

The F-statistic for testing $H_0 : \delta_1 = 0$ and $\delta_2 = 0$ is

$$F_{\delta_1=0,\delta_2=0} = \frac{(0.301558 - 0.288379)/2}{0.288379/116} = 2.651$$

The 95% quantile, $F_{0.95}(2, 116)$, is approximately 3.07. Since $F_{\delta_1=0,\delta_2=0} = 2.651 < 3.07 = F_{0.95}(2, 116)$ we do not reject $H_0 : \delta_1 = 0$ and $\delta_2 = 0$ at the 5% significance level. It is interesting to note that when we allow both α and β to differ in the two subsamples we cannot reject the hypothesis that these parameters are the same between two samples but if we only allow β to differ between the two samples we can reject the hypothesis that β is the same.

2.5. Other types of Structural Change in β

An interesting question regarding the beta of an asset concerns the stability of beta over the market cycle. For example, consider the following situations. Suppose that the beta of an asset is greater than 1 if the market is in an "up cycle", $R_{Mt} - r_f > 0$, and less than 1 in a "down cycle", $R_{Mt} - r_f < 0$. This would be a very desirable asset to hold since it accentuates positive market movements but down plays negative market movements. We can investigate this hypothesis using a dummy variable as follows. Define

$$\begin{aligned} D_t^{up} &= 1, \; R_{Mt} - r_f > 0 \\ &= 0, \; R_{Mt} - r_f \leq 0. \end{aligned}$$

Then D_t^{up} divides the sample into "up market" movements and "down market" movements. The regression that allows beta to differ depending on the market cycle is then

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \delta D_t^{up} \cdot (R_{Mt} - r_f) + \varepsilon_t.$$

In the down cycle, when $D_t^{up} = 0$, the model is

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t$$

and β captures the down market beta, and in the up market, when $D_t^{up} = 1$, the model is

$$R_t - r_f = \alpha + (\beta + \delta)(R_{Mt} - r_f) + \varepsilon_t$$

so that $\beta + \delta$ capture the up market beta. The hypothesis that β does not vary over the market cycle is

$$H_0 : \delta = 0 \text{ vs. } H_1 : \delta \neq 0 \tag{2.2}$$

and can be tested with the simple t-statistic $t_{\delta=0} = \frac{\widehat{\delta}-0}{\widehat{SE}(\widehat{\delta})}$.

If the estimated value of δ is found to be statistically greater than zero we might then want to go on to test the hypothesis that the up market beta is greater than one. Since the up market beta is equal to $\beta + \delta$ this corresponds to testing

$$H_0 : \beta + \delta \leq 1 \text{ vs. } H_1 : \beta + \delta \geq 1$$

which can be tested using the t-statistic

$$t_{\beta+\delta=1} = \frac{\widehat{\beta} + \widehat{\delta} - 1}{\widehat{SE}(\widehat{\beta} + \widehat{\delta})}.$$

Since this is a one-sided test we will reject the null hypothesis at the 5% level if $t_{\beta+\delta=1} < -t_{0.05,T-3}$.

Example 2.6. *CAPM regression for IBM and DEC*

For IBM the CAPM regression allowing β to vary over the market cycle (1978.01 - 1982.12) is

$$\begin{aligned} \widehat{R_{IBM,t}} - r_f &= \underset{(0.0111)}{-0.0019} + \underset{(0.1476)}{0.3163} \cdot (R_{Mt} - r_f) + \underset{(0.2860)}{0.0552} \cdot D_t^{up} \cdot (R_{Mt} - r_f) \\ R^2 &= 0.201, \widehat{\sigma} = 0.053 \end{aligned}$$

Notice that $\widehat{\delta} = 0.0552$, with a standard error of 0.2860, is close to zero and not estimated very precisely. Consequently, $t_{\delta=0} = \frac{0.0552}{0.2860} = 0.1929$ is not significant at any reasonable significance level and we therefore reject the hypothesis that beta varies over the market cycle. However, the results are very different for DEC (Digital Electronics):

$$\begin{aligned} \widehat{R_{DEC,t}} - r_f &= \underset{(0.0134)}{-0.0248} + \underset{(0.1779)}{0.3689} \cdot (R_{Mt} - r_f) + \underset{(0.3446)}{0.8227} \cdot D_t^{up} \cdot (R_{Mt} - r_f) \\ R^2 &= 0.460, \widehat{\sigma} = 0.064. \end{aligned}$$

Here $\widehat{\delta} = 0.8227$, with a standard error of 0.3446, is statistically different from zero at the 5% level since $t_{\delta=0} = 2.388$. The estimate of the down market beta is 0.3689, which is less than one, and the up market beta is $0.3689 + 0.8227 = 1.1916$, which is greater than one. The estimated standard error for $\widehat{\beta}+\widehat{\delta}$ requires the estimated variances of $\widehat{\beta}$ and $\widehat{\delta}$ and the estimated covariance between $\widehat{\beta}$ and $\widehat{\delta}$ (which can be obtained from Eviews) and is given by

$$\begin{aligned}
\widehat{var}(\widehat{\beta}+\widehat{\delta}) &= \widehat{var}(\widehat{\beta}) + \widehat{var}(\widehat{\delta}) + 2\cdot\widehat{cov}(\widehat{\beta},\widehat{\delta}) \\
&= 0.031634 + 0.118656 + 2\cdot -0.048717 \\
&= 0.052856, \\
\widehat{SE}(\widehat{\beta}+\widehat{\delta}) &= \sqrt{\widehat{var}(\widehat{\beta}+\widehat{\delta})} = \sqrt{0.052856} = 0.2299
\end{aligned}$$

Then $t_{\beta+\delta=1} = \frac{1.1916-1}{0.2299} = 0.8334$ which is less than $t_{0.05,57} = 1.65$ so we do not reject the hypothesis that the up market beta is less than or equal to one.

Introduction to Financial Econometrics
Hypothesis Testing in the Market Model

Eric Zivot
Department of Economics
University of Washington

February 29, 2000

1 Hypothesis Testing in the Market Model

In this chapter, we illustrate how to carry out some simple hypothesis tests concerning the parameters of the excess returns market model regression.

1.1 A Review of Hypothesis Testing Concepts

To be completed.

1.2 Testing the Restriction $\alpha = 0$.

Using the market model regression,

$$\begin{aligned} R_t &= \alpha + \beta R_{Mt} + \varepsilon_t,\ t = 1, ..., T \\ \varepsilon_t &\sim iid\ N(0, \sigma_\varepsilon^2),\ \varepsilon_t \text{ is independent of } R_{Mt} \end{aligned} \tag{1}$$

consider testing the null or maintained hypothesis $\alpha = 0$ against the alternative that $\alpha \neq 0$

$$H_0 : \alpha = 0\ vs.\ H_1 : \alpha \neq 0.$$

If H_0 is true then the market model regression becomes

$$R_t = \beta R_{Mt} + \varepsilon_t$$

and $E[R_t|R_{Mt} = r_{Mt}] = \beta r_{Mt}$. We will reject the null hypothesis, $H_0 : \alpha = 0$, if the estimated value of α is either much larger than zero or much smaller than zero. Assuming $H_0 : \alpha = 0$ is true, $\hat{\alpha} \sim N(0, SE(\hat{\alpha})^2)$ and so is fairly unlikely that $\hat{\alpha}$ will be more than 2 values of $SE(\hat{\alpha})$ from zero. To determine how big the estimated value of α needs to be in order to reject the null hypothesis we use the t-statistic

$$t_{\alpha=0} = \frac{\widehat{\alpha} - 0}{\widehat{SE}(\widehat{\alpha})},$$

where $\widehat{\alpha}$ is the least squares estimate of α and $\widehat{SE}(\widehat{\alpha})$ is its estimated standard error. The value of the t-statistic, $t_{\alpha=0}$, gives the number of estimated standard errors that $\widehat{\alpha}$ is from zero. If the absolute value of $t_{\alpha=0}$ is much larger than 2 then the data cast considerable doubt on the null hypothesis $\alpha = 0$ whereas if it is less than 2 the data are in support of the null hypothesis[1]. To determine how big $\mid t_{\alpha=0}\mid$ needs to be to reject the null, we use the fact that under the statistical assumptions of the market model and *assuming the null hypothesis is true*

$$t_{\alpha=0} \sim Student - t \text{ with } T - 2 \text{ degrees of freedom}$$

If we set the significance level (the probability that we reject the null given that the null is true) of our test at, say, 5% then our decision rule is

$$\text{Reject } H_0 : \alpha = 0 \text{ at the 5\% level if } |t_{\alpha=0}| > t_{T-2}(0.025)$$

where t_{T-2} is the $2\frac{1}{2}\%$ critical value from a Student-t distribution with $T - 2$ degrees of freedom.

Example 1 *Market Model Regression for IBM*

Consider the estimated MM regression equation for IBM using monthly data from January 1978 through December 1982:

$$\widehat{R}_{IBM,t} = \underset{(0.0068)}{-0.0002} + \underset{(0.0888)}{0.3390} \cdot R_{Mt},\ R^2 = 0.20,\ \widehat{\sigma}_{\varepsilon} = 0.0524$$

where the estimated standard errors are in parentheses. Here $\widehat{\alpha} = -0.0002$, which is very close to zero, and the estimated standard error, $\widehat{SE}(\hat{\alpha}) = 0.0068$, is much larger than $\widehat{\alpha}$. The t-statistic for testing $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$ is

$$t_{\alpha=0} = \frac{-0.0002 - 0}{0.0068} = -0.0363$$

so that $\widehat{\alpha}$ is only 0.0363 estimated standard errors from zero. Using a 5% significance level, $t_{58}(0.025) \approx 2$ and

$$|t_{\alpha=0}| = 0.0363 < 2$$

so we do not reject $H_0 : \alpha = 0$ at the 5% level.

[1] This interpretation of the t-statistic relies on the fact that, assuming the null hypothesis is true so that $\alpha = 0$, $\widehat{\alpha}$ is normally distributed with mean 0 and estimated variance $\widehat{SE}(\widehat{\alpha})^2$.

1.3 Testing Hypotheses about β

In the market model regression β measures the contribution of an asset to the variability of the market index portfolio. One hypothesis of interest is to test if the asset has the same level of risk as the market index against the alternative that the risk is different from the market:

$$H_0 : \beta = 1 \ vs. \ H_1 : \beta \neq 1.$$

The data cast doubt on this hypothesis if the estimated value of β is much different from one. This hypothesis can be tested using the t-statistic

$$t_{\beta=1} = \frac{\widehat{\beta} - 1}{\widehat{SE}(\widehat{\beta})}$$

which measures how many estimated standard errors the least squares estimate of β is from one. The null hypothesis is reject at the 5% level, say, if $|t_{\beta=1}| > t_{T-2}(0.025)$. Notice that this is a two-sided test.

Alternatively, one might want to test the hypothesis that the risk of an asset is strictly less than the risk of the market index against the alternative that the risk is greater than or equal to that of the market:

$$H_0 : \beta = 1 \ vs. \ H_1 : \beta \geq 1.$$

Notice that this is a one-sided test. We will reject the null hypothesis only if the estimated value of β much greater than one. The t-statistic for testing this null hypothesis is the same as before but the decision rule is different. Now we reject the null at the 5% level if

$$t_{\beta=1} < -t_{T-2}(0.05)$$

where $t_{T-2}(0.05)$ is the one-sided 5% critical value of the Student-t distribution with $T-2$ degrees of freedom.

Example 2 *MM Regression for IBM cont'd*

Continuing with the previous example, consider testing $H_0 : \beta = 1 \ vs. \ H_1 : \beta \neq 1$. Notice that the estimated value of β is 0.3390, with an estimated standard error of 0.0888, and is fairly far from the hypothesized value $\beta = 1$. The t-statistic for testing $\beta = 1$ is

$$t_{\beta=1} = \frac{0.3390 - 1}{0.0888} = -7.444$$

which tells us that $\widehat{\beta}$ is more than 7 estimated standard errors below one. Since $t_{0.025,58} \approx 2$ we easily reject the hypothesis that $\beta = 1$.

Now consider testing $H_0 : \beta = 1 \ vs. \ H_1 : \beta \geq 1$. The t-statistic is still -7.444 but the critical value used for the test is now $-t_{58}(0.05) \approx -1.671$. Clearly, $t_{\beta=1} = -7.444 < -1.671 = -t_{58}(0.05)$ so we reject this hypothesis.

1.4 Testing Joint Hypotheses about α and β

Often it is of interest to formulate hypothesis tests that involve both α and β. For example, consider the joint hypothesis that $\alpha = 0$ and $\beta = 1$:

$$H_0 : \alpha = 0 \text{ and } \beta = 1.$$

The null will be rejected if either $\alpha \neq 0, \beta \neq 1$ or both.. Thus the alternative is formulated as

$$H_1 : \alpha \neq 0, \text{or } \beta \neq 1 \text{ or } \alpha \neq 0 \text{ and } \beta \neq 1.$$

This type of joint hypothesis is easily tested using a so-called *F-test*. The idea behind the F-test is to estimate the model imposing the restrictions specified under the null hypothesis and compare the fit of the restricted model to the fit of the model with no restrictions imposed.

The fit of the unrestricted (UR) excess return market model is measured by the (unrestricted) sum of squared residuals (RSS)

$$SSR_{UR} = SSR(\hat{\alpha}, \hat{\beta}) = \sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \sum_{t=1}^{T} (R_t - \hat{\alpha} - \hat{\beta} R_{Mt})^2.$$

Recall, this is the quantity that is minimized during the least squares algorithm. Now, the market model imposing the restrictions under H_0 is

$$\begin{aligned} R_t &= 0 + 1 \cdot (R_{Mt} - r_f) + \varepsilon_t \\ &= R_{Mt} + \varepsilon_t. \end{aligned}$$

Notice that there are no parameters to be estimated in this model which can be seen by subtracting R_{Mt} from both sides of the restricted model to give

$$R_t - R_{Mt} = \tilde{\varepsilon}_t$$

The fit of the restricted (R) model is then measured by the restricted sum of squared residuals

$$SSR_R = SSR(\alpha = 0, \beta = 1) = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2 = \sum_{t=1}^{T} (R_t - R_{Mt})^2.$$

Now since the least squares algorithm works to minimize SSR, the restricted error sum of squares, SSR_R, must be at least as big as the unrestricted error sum of squares, SSR_{UR}. If the restrictions imposed under the null are true then $SSR_R \approx SSR_{UR}$ (with SSR_R always slightly bigger than SSR_{UR}) but if the restrictions are not true then SSR_R will be quite a bit bigger than SSR_{UR}. The *F-statistic* measures the (adjusted) percentage difference in fit between the restricted and unrestricted models and is given by

$$F = \frac{(SSR_R - SSR_{UR})/q}{SSR_{UR}/(T-k)} = \frac{(SSR_R - SSR_{UR})}{q \cdot \hat{\sigma}^2_{\varepsilon,UR}},$$

where q equals the number of restrictions imposed under the null hypothesis, k denotes the number of regression coefficients estimated under the unrestricted model and $\widehat{\sigma}^2_{\varepsilon,UR}$ denotes the estimated variance of ε_t under the unrestricted model. Under the assumption that the null hypothesis is true, the F-statistic is distributed as an F random variable with q and $T-2$ degrees of freedom:

$$F \sim F_{q,T-2}.$$

Notice that an F random variable is always positive since $SSR_R > SSR_{UR}$. The null hypothesis is rejected, say at the 5% significance level, if

$$F > F_{q,T-k}(0.05)$$

where $F_{q,T-k}(0.05)$ is the 95% quantile of the distribution of $F_{q,T-k}$.

For the hypothesis $H_0 : \alpha = 0$ and $\beta = 1$ there are $q = 2$ restrictions under the null and $k = 2$ regression coefficients estimated under the unrestricted model. The F-statistic is then

$$F_{\alpha=0,\beta=1} = \frac{(SSR_R - SSR_{UR})/2}{SSR_{UR}/(T-2)}$$

Example 3 *MM Regression for IBM cont'd*

Consider testing the hypothesis $H_0 : \alpha = 0$ and $\beta = 1$ for the IBM data. The unrestricted error sum of squares, SSR_{UR}, is obtained from the unrestricted regression output in figure 2 and is called **Sum Square Resid:**

$$SSR_{UR} = 0.159180.$$

To form the restricted sum of squared residuals, we create the new variable $\widetilde{\varepsilon}_t = R_t - R_{Mt}$ and form the sum of squares $SSR_R = \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2 = 0.31476$. Notice that $SSR_R > SSR_{UR}$. The F-statistic is then

$$F_{\alpha=0,\beta=1} = \frac{(0.31476 - 0.159180)/2}{0.159180/58} = 28.34.$$

The 95% quantile of the F-distribution with 2 and 58 degrees of freedom is about 3.15. Since $F_{\alpha=0,\beta=1} = 28.34 > 3.15 = F_{2,58}(0.05)$ we reject $H_0 : \alpha = 0$ and $\beta = 1$ at the 5% level.

1.5 Testing the Stability of α and β over time

In many applications of the MM, α and β are estimated using past data and the estimated values of α and β are used to make decision about asset allocation and risk over some future time period. In order for this analysis to be useful, it is assumed that the unknown values of α and β are constant over time. Since the risk characteristics of assets may change over time it is of interest to know if α and β change over time. To illustrate, suppose we have a ten year sample of monthly data ($T = 120$) on returns that we split into two five year subsamples. Denote the first five years as $t = 1, ..., T_B$ and the second five years as $t = T_{B+1}, ..., T$. The date $t = T_B$ is the "break date" of the sample and it is chosen arbitrarily in this context. Since the samples are of equal size (although they do not have to be) $T - T_B = T_B$ or $T = 2 \cdot T_B$. The market model regression which assumes that both α and β are constant over the entire sample is

$$\begin{aligned} R_t &= \alpha + \beta R_{Mt} + \varepsilon_t, \ t = 1, \ldots, T \\ \varepsilon_t &\sim iid\ N(0, \sigma^2) \text{ independent of } R_{Mt}. \end{aligned}$$

There are three main cases of interest: (1) β may differ over the two subsamples; (2) α may differ over the two subsamples; (3) α and β may differ over the two subsamples.

1.5.1 Testing Structural Change in β only

If α is the same but β is different over the subsamples then we really have two market model regressions

$$\begin{aligned} R_t &= \alpha + \beta_1 R_{Mt} + \varepsilon_t, \ t = 1, \ldots, T_B \\ R_t &= \alpha + \beta_2 R_{Mt} + \varepsilon_t, \ t = T_{B+1}, \ldots, T \end{aligned}$$

that share the same intercept α but have different slopes $\beta_1 \neq \beta_2$. We can capture such a model very easily using a *step dummy variable* defined as

$$\begin{aligned} D_t &= 0, \ t \leq T_B \\ &= 1, \ t > T_B \end{aligned}$$

and re-writing the MM regression as the multiple regression

$$R_t = \alpha + \beta R_{Mt} + D_t R_{Mt} + \varepsilon_t.$$

The model for the first subsample when $D_t = 0$ is

$$R_t = \alpha + \beta R_{Mt} + \varepsilon_t, \ t = 1, \ldots, T_B$$

and the model for the second subsample when $D_t = 1$ is

$$\begin{aligned} R_t &= \alpha + \beta R_{Mt} + \delta R_{Mt} + \varepsilon_t, \ t = T_{B+1}, \ldots, T \\ &= \alpha + (\beta + \delta) R_{Mt} + \varepsilon_t. \end{aligned}$$

Notice that the "beta" in the first sample is $\beta_1 = \beta$ and the beta in the second subsample is $\beta_2 = \beta + \delta$. If $\delta < 0$ the second sample beta is smaller than the first sample beta and if $\delta > 0$ the beta is larger.

We can test the constancy of beta over time by testing $\delta = 0$:

H_0 : (beta is constant over two sub-samples) $\delta = 0$ vs. H_1 : (beta is not constant over two sub-samples

The test statistic is simply the t-statistic

$$t_{\delta=0} = \frac{\widehat{\delta} - 0}{\widehat{SE}(\widehat{\delta})} = \frac{\widehat{\delta}}{\widehat{SE}(\widehat{\delta})}$$

and we reject the hypothesis $\delta = 0$ at the 5% level, say, if $|t_{\delta=0}| > t_{T-3}(0.025)$.

Example 4 *MM regression for IBM cont'd*

Consider the estimated MM regression equation for IBM using ten years of monthly data from January 1978 through December 1987. We want to know if the beta on IBM is using the first five years of data (January 1978 - December 1982) is different from the beta on IBM using the second five years of data (January 1983 - December 1987). We define the step dummy variable

$$\begin{aligned} D_t &= 1 \text{ if } t > \text{December 1982} \\ &= 0, \text{ otherwise} \end{aligned}$$

The estimated (unrestricted) model allowing for structural change in β is given by

$$\begin{aligned} \widehat{R_{IBM,t}} &= \underset{(0.0045)}{-0.0001} + \underset{(0.0837)}{0.3388} \cdot R_{M,t} + \underset{(0.1366)}{0.3158} \cdot D_t \cdot R_{M,t}, \\ R^2 &= 0.311, \ \widehat{\sigma}_\varepsilon = 0.0496. \end{aligned}$$

The estimated value of β is 0.3388, with a standard error of 0.0837, and the estimated value of δ is 0.3158, with a standard error of 0.1366. The t-statistic for testing $\delta = 0$ is given by

$$t_{\delta=0} = \frac{0.3158}{0.1366} = 2.312$$

which is greater than $t_{117}(0.025) = 1.98$ so we reject the null hypothesis (at the 5% significance level) that beta is the same over the two subsamples. The implied estimate of beta over the period January 1983 - December 1987 is

$$\hat{\beta} + \hat{\delta} = 0.3388 + 0.3158 = 0.6546.$$

It appears that IBM has become more risky.

1.5.2 Testing Structural Change in α and β

Now consider the case where both α and β are allowed to be different over the two subsamples:

$$
\begin{aligned}
R_t &= \alpha_1 + \beta_1 R_{Mt} + \varepsilon_t,\ t = 1, \ldots, T_B \\
R_t &= \alpha_2 + \beta_2 R_{Mt} + \varepsilon_t,\ t = T_{B+1}, \ldots, T
\end{aligned}
$$

The dummy variable specification in this case is

$$
R_t = \alpha + \beta R_{Mt} + \delta_1 \cdot D_t + \delta_2 \cdot D_t R_{Mt} + \varepsilon_t,\ t = 1, \ldots, T.
$$

When $D_t = 0$ the model becomes

$$
R_t = \alpha + \beta R_{Mt} + \varepsilon_t,\ t = 1, \ldots, T_B,
$$

so that $\alpha_1 = \alpha$ and $\beta_1 = \beta$, and when $D_t = 1$ the model is

$$
R_t = (\alpha + \delta_1) + (\beta + \delta_2) R_{Mt} + \varepsilon_t,\ t = T_{B+1}, \ldots, T,
$$

so that $\alpha_2 = \alpha + \delta_1$ and $\beta_2 = \beta + \delta_2$. The hypothesis of no structural change is now

$$
H_0 : \delta_1 = 0 \text{ and } \delta_2 = 0 \text{ vs. } H_1 : \delta_1 \neq 0 \text{ or } \delta_2 \neq 0 \text{ or } \delta_1 \neq 0 \text{ and } \delta_2 \neq 0.
$$

The test statistic for this joint hypothesis is the F-statistic

$$
F_{\delta_1 = 0, \delta_2 = 0} = \frac{(SSR_R - SSR_{UR})/2}{SSR_{UR}/(T-4)}
$$

since there are two restrictions and four regression parameters estimated under the unrestricted model. The unrestricted (UR) model is the dummy variable regression that allows the intercepts and slopes to differ in the two subsamples and the restricted model (R) is the regression where these parameters are constrained to be the same in the two subsamples.

The unrestricted error sum of squares, SSR_{UR}, can be computed in two ways. The first way is based on the dummy variable regression. The second is based on estimating separate regression equations for the two subsamples and adding together the resulting error sum of squares. Let SSR_1 and SSR_2 denote the error sum of squares from separate regressions. Then

$$
SSR_{UR} = SSR_1 + SSR_2.
$$

Example 5 *MM regression for IBM cont'd*

The unrestricted regression is

$$\begin{aligned}\widehat{R_{IBM,t}} &= \underset{(0.0065)}{-0.0001} + \underset{(0.0845)}{0.3388} \cdot R_{Mt} \\ &\quad + \underset{(0.0092)}{0.0002} \cdot D_t + \underset{(0.1377)}{0.3158} \cdot D_t \cdot R_{Mt}, \\ R^2 &= 0.311,\ \widehat{\sigma}_\varepsilon = 0.050,\ SSR_{UR} = 0.288379,\end{aligned}$$

and the restricted regression is

$$\begin{aligned}\widehat{R_{IBM,t}} &= \underset{(0.0046)}{-0.0005} + \underset{(0.0675)}{0.4568} \cdot R_{Mt}, \\ R^2 &= 0.279,\ \widehat{\sigma}_\varepsilon = 0.051,\ SSR_R = 0.301558.\end{aligned}$$

The F-statistic for testing $H_0 : \delta_1 = 0$ and $\delta_2 = 0$ is

$$F_{\delta_1=0,\delta_2=0} = \frac{(0.301558 - 0.288379)/2}{0.288379/116} = 2.651$$

The 95% quantile, $F_{2,116}(0.05)$, is approximately 3.07. Since $F_{\delta_1=0,\delta_2=0} = 2.651 < 3.07 = F_{2,116}(0.05)$ we do not reject $H_0 : \delta_1 = 0$ and $\delta_2 = 0$ at the 5% significance level. It is interesting to note that when we allow both α and β to differ in the two subsamples we cannot reject the hypothesis that these parameters are the same between two samples but if we only allow β to differ between the two samples we can reject the hypothesis that β is the same.

1.6 Other types of Structural Change in β

An interesting question regarding the beta of an asset concerns the stability of beta over the market cycle. For example, consider the following situations. Suppose that the beta of an asset is greater than 1 if the market is in an "up cycle", $R_{Mt} > 0$, and less than 1 in a "down cycle", $R_{Mt} < 0$. This would be a very desirable asset to hold since it accentuates positive market movements but down plays negative market movements. We can investigate this hypothesis using a dummy variable as follows. Define

$$\begin{aligned}D_t^{up} &= 1,\ R_{Mt} > 0 \\ &= 0,\ R_{Mt} \leq 0.\end{aligned}$$

Then D_t^{up} divides the sample into "up market" movements and "down market" movements. The regression that allows beta to differ depending on the market cycle is then

$$R_t = \alpha + \beta R_{Mt} + \delta D_t^{up} \cdot R_{Mt} + \varepsilon_t.$$

In the down cycle, when $D_t^{up} = 0$, the model is

$$R_t = \alpha + \beta R_{Mt} + \varepsilon_t$$

and β captures the down market beta, and in the up market, when $D_t^{up} = 1$, the model is

$$R_t = \alpha + (\beta + \delta)R_{Mt} + \varepsilon_t$$

so that $\beta + \delta$ capture the up market beta. The hypothesis that β does not vary over the market cycle is

$$H_0 : \delta = 0 \text{ vs. } H_1 : \delta \neq 0 \tag{2}$$

and can be tested with the simple t-statistic $t_{\delta=0} = \frac{\widehat{\delta}-0}{\widehat{SE}(\widehat{\delta})}$.

If the estimated value of δ is found to be statistically greater than zero we might then want to go on to test the hypothesis that the up market beta is greater than one. Since the up market beta is equal to $\beta + \delta$ this corresponds to testing

$$H_0 : \beta + \delta = 1 \text{ vs. } H_1 : \beta + \delta \geq 1$$

which can be tested using the t-statistic

$$t_{\beta+\delta=1} = \frac{\widehat{\beta} + \widehat{\delta} - 1}{\widehat{SE}(\widehat{\beta} + \widehat{\delta})}.$$

Since this is a one-sided test we will reject the null hypothesis at the 5% level if $t_{\beta+\delta=1} < -t_{0.05,T-3}$.

Example 6 *MM regression for IBM and DEC*

For IBM the CAPM regression allowing β to vary over the market cycle (1978.01 - 1982.12) is

$$\begin{aligned}
\widehat{R_{IBM,t}} &= \underset{(0.0111)}{-0.0019} + \underset{(0.1476)}{0.3163} \cdot R_{Mt} + \underset{(0.2860)}{0.0552} \cdot D_t^{up} \cdot R_{Mt} \\
R^2 &= 0.201, \widehat{\sigma} = 0.053
\end{aligned}$$

Notice that $\widehat{\delta} = 0.0552$, with a standard error of 0.2860, is close to zero and not estimated very precisely. Consequently, $t_{\delta=0} = \frac{0.0552}{0.2860} = 0.1929$ is not significant at any reasonable significance level and we therefore reject the hypothesis that beta varies over the market cycle. However, the results are very different for DEC (Digital Electronics):

$$\begin{aligned}
\widehat{R_{DEC,t}} &= \underset{(0.0134)}{-0.0248} + \underset{(0.1779)}{0.3689} \cdot R_{Mt} + \underset{(0.3446)}{0.8227} \cdot D_t^{up} \cdot R_{Mt} \\
R^2 &= 0.460, \widehat{\sigma} = 0.064.
\end{aligned}$$

Here $\widehat{\delta} = 0.8227$, with a standard error of 0.3446, is statistically different from zero at the 5% level since $t_{\delta=0} = 2.388$. The estimate of the down market beta is 0.3689, which is less than one, and the up market beta is 0.3689 + 0.8227 = 1.1916, which is greater than one. The estimated standard error for $\widehat{\beta} + \widehat{\delta}$ requires the estimated variances of $\widehat{\beta}$ and $\widehat{\delta}$ and the estimated covariance between $\widehat{\beta}$ and $\widehat{\delta}$ and is given by

$$\begin{aligned}
\widehat{var}(\widehat{\beta} + \widehat{\delta}) &= \widehat{var}(\widehat{\beta}) + \widehat{var}(\widehat{\delta}) + 2 \cdot \widehat{cov}(\widehat{\beta}, \widehat{\delta}) \\
&= 0.031634 + 0.118656 + 2 \cdot -0.048717 \\
&= 0.052856, \\
\widehat{SE}(\widehat{\beta} + \widehat{\delta}) &= \sqrt{\widehat{var}(\widehat{\beta} + \widehat{\delta})} = \sqrt{0.052856} = 0.2299
\end{aligned}$$

Then $t_{\beta+\delta=1} = \frac{1.1916-1}{0.2299} = 0.8334$ which is less than $t_{0.05,57} = 1.65$ so we do not reject the hypothesis that the up market beta is less than or equal to one.

Chapter 1

The Constant Expected Return Model

The first model of asset returns we consider is the very simple *constant expected return* (CER) model. This model assumes that an asset's return over time is normally distributed with a constant (time invariant) mean and variance The model also assumes that the correlations between asset returns are constant over time. Although this model is very simple, it allows us to discuss and develop several important econometric topics such as estimation, hypothesis testing, forecasting and model evaluation.

1.0.1 Constant Expected Return Model Assumptions

Let R_{it} denote the continuously compounded return on an asset i at time t. We make the following assumptions regarding the probability distribution of R_{it} for $i = 1, \ldots, N$ assets over the time horizon $t = 1, \ldots, T$.

1. Normality of returns: $R_{it} \sim N(\mu_i, \sigma_i^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

2. Constant variances and covariances: $cov(R_{it}, R_{jt}) = \sigma_{ij}$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

3. No serial correlation across assets over time: $cov(R_{it}, R_{js}) = 0$ for $t \neq s$ and $i, j = 1, \ldots, N$.

Assumption 1 states that in every time period asset returns are normally distributed and that the mean and the variance of each asset return is constant over time. In particular, we have for each asset i and every time period

t

$$
\begin{aligned}
E[R_{it}] &= \mu_i \\
var(R_{it}) &= \sigma_i^2
\end{aligned}
$$

The second assumption states that the contemporaneous covariances between assets are constant over time. Given assumption 1, assumption 2 implies that the contemporaneous correlations between assets are constant over time as well. That is, for all assets and time periods

$$corr(R_{it}, R_{jt}) = \rho_{ij}$$

The third assumption stipulates that all of the asset returns are uncorrelated over time[1]. In particular, for a given asset i the returns on the asset are *serially uncorrelated* which implies that

$$corr(R_{it}, R_{is}) = cov(R_{it}, R_{is}) = 0 \text{ for all } t \neq s.$$

Additionally, the returns on all possible pairs of assets i and j are serially uncorrelated which implies that

$$corr(R_{it}, R_{js}) = cov(R_{it}, R_{js}) = 0 \text{ for all } i \neq j \text{ and } t \neq s.$$

Assumptions 1-3 indicate that all asset returns at a given point in time are jointly (multivariate) normally distributed and that this joint distribution stays constant over time. Clearly these are very strong assumptions. However, they allow us to development a straightforward probabilistic model for asset returns as well as statistical tools for estimating the parameters of the model and testing hypotheses about the parameter values and assumptions.

1.0.2 Regression Model Representation

A convenient mathematical representation or *model* of asset returns can be given based on assumptions 1-3. This is the *constant expected return* (CER) *regression* model. For assets $i = 1, \ldots, N$ and time periods $t = 1, \ldots, T$ the CER model is represented as

$$
\begin{aligned}
R_{it} &= \mu_i + \varepsilon_{it} && (1.1) \\
\varepsilon_{it} &\sim iid.\ N(0, \sigma_i^2) \\
cov(\varepsilon_{it}, \varepsilon_{jt}) &= \sigma_{ij} && (1.2)
\end{aligned}
$$

[1]Since all assets are assumed to be normally distributed (assumption 1), uncorrelatedness implies the stronger condition of independence.

where μ_i is a constant and ε_{it} is a normally distributed random variable with mean zero and variance σ_i^2 . Notice that the random error term ε_{it} is independent of ε_{js} for all time periods $t \neq s$. The notation $\varepsilon_{it} \sim iid.\ N(0, \sigma_i^2)$ stipulates that the random variable ε_{it} is serially independent and identically distributed as a normal random variable with mean zero and variance σ_i^2. This implies that, $E[\varepsilon_{it}] = 0$, $var(\varepsilon_{it}) = \sigma_i^2$ and $cov(\varepsilon_{it}, \varepsilon_{js}) = 0$ for $i \neq j$ and $t \neq s$.

Using the basic properties of expectation, variance and covariance discussed in chapter 2, we can derive the following properties of returns. For expected returns we have

$$E[R_{it}] = E[\mu_i + \varepsilon_{it}] = \mu_i + E[\varepsilon_{it}] = \mu_i,$$

since μ_i is constant and $E[\varepsilon_{it}] = 0$. Regarding the variance of returns, we have

$$var(R_{it}) = var(\mu_i + \varepsilon_{it}) = var(\varepsilon_{it}) = \sigma_i^2$$

which uses the fact that the variance of a constant (μ_i) is zero. For covariances of returns, we have

$$cov(R_{it}, R_{jt}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{jt}) = cov(\varepsilon_{it}, \varepsilon_{jt}) = \sigma_{ij}$$

and

$$cov(R_{it}, R_{js}) = cov(\mu_i + \varepsilon_{it}, \mu_j + \varepsilon_{js}) = cov(\varepsilon_{it}, \varepsilon_{js}) = 0,\ t \neq s,$$

which use the fact that adding constants to two random variables does not affect the covariance between them. Given that covariances and variances of returns are constant over time gives the result that correlations between returns over time are also constant:

$$\begin{aligned} corr(R_{it}, R_{jt}) &= \frac{cov(R_{it}, R_{jt})}{\sqrt{var(R_{it})var(R_{jt})}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij}, \\ corr(R_{it}, R_{js}) &= \frac{cov(R_{it}, R_{js})}{\sqrt{var(R_{it})var(R_{js})}} = \frac{0}{\sigma_i \sigma_j} = 0,\ i \neq j, t \neq s. \end{aligned}$$

Finally, since the random variable ε_{it} is independent and identically distributed ($i.i.d.$) normal the asset return R_{it} will also be $i.i.d.$ normal:

$$R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2).$$

Hence, the CER model (1.1) for R_{it} is equivalent to the model implied by assumptions 1-3.

1.0.3 Interpretation of the CER Regression Model

The CER model has a very simple form and is identical to the *measurement error model* in the statistics literature. In words, the model states that each asset return is equal to a constant μ_i (the expected return) plus a normally distributed random variable ε_{it} with mean zero and constant variance. The random variable ε_{it} can be interpreted as representing the *unexpected news* concerning the value of the asset that arrives between times $t-1$ and time t. To see this, note that using (1.1) we can write ε_{it} as

$$\begin{aligned} \varepsilon_{it} &= R_{it} - \mu_i \\ &= R_{it} - E[R_{it}] \end{aligned}$$

so that ε_{it} is defined to be the deviation of the random return from its expected value. If the news between times $t-1$ and time t is good, then the realized value of ε_{it} is positive and the observed return is above its expected value μ_i. If the news is bad, then ε_{jt} is negative and the observed return is less than expected. The assumption that $E[\varepsilon_{it}] = 0$ means that news, on average, is neutral; neither good nor bad. The assumption that $var(\varepsilon_{it}) = \sigma_i^2$ can be interpreted as saying that volatility of news arrival is constant over time. The random news variable affecting asset i, ε_{it}, is allowed to be contemporaneously correlated with the random news variable affecting asset j, ε_{jt}, to capture the idea that news about one asset may spill over and affect another asset. For example, let asset i be Microsoft and asset j be Apple Computer. Then one interpretation of news in this context is general news about the computer industry and technology. Good news should lead to positive values of ε_{it} and ε_{jt}. Hence these variables will be positively correlated.

Time Aggregation and the CER Model

The CER model with continuously compounded returns has the following nice property with respect to the interpretation of ε_{it} as news. Consider the default case where R_{it} is interpreted as the continuously compounded monthly return on asset i. Suppose we are interested in the annual continuously compounded return $R_{it}^A = R_{it}(12)$? Since multiperiod continuously compounded returns are additive, $R_{it}(12)$ is the sum of 12 monthly continuously compounded returns[2]:

$$R_{it}^{A} = R_{it}(12) = \sum_{t=0}^{11} R_{it-k} = R_{it} + R_{it-1} + \cdots + R_{it-11}$$

Using the CER model representation (1.1) for the monthly return R_{it} we may express the annual return $R_{it}(12)$ as

$$\begin{aligned} R_{it}(12) &= \sum_{t=0}^{11}(\mu_i + \varepsilon_{it}) \\ &= 12 \cdot \mu_i + \sum_{t=0}^{11} \varepsilon_{it} \\ &= \mu_i^A + \varepsilon_{it}^A \end{aligned}$$

where $\mu_i^A = 12 \cdot \mu_i$ is the annual expected return on asset i and $\varepsilon_{it}^A = \sum_{k=0}^{11} \varepsilon_{it-k}$ is the annual random news component. Hence, the annual expected return, μ_i^A, is simply 12 times the monthly expected return, μ_i. The annual random news component, ε_{it}^A , is the accumulation of news over the year. Using the results from chapter 2 about the variance of a sum of random variables, the variance of the annual news component is just 12 time the variance of the monthly new component:

$$\begin{aligned} var(\varepsilon_{it}^A) &= var\left(\sum_{k=0}^{11} \varepsilon_{it-k})\right) \\ &= \sum_{k=0}^{11} var(\varepsilon_{it-k}) \text{ since } \varepsilon_{it} \text{ is uncorrelated over time} \\ &= \sum_{k=0}^{11} \sigma_i^2 \text{ since } var(\varepsilon_{it}) \text{ is constant over time} \\ &= 12 \cdot \sigma_i^2 \\ &= var(R_{it}^A) \end{aligned}$$

[2]For simplicity of exposition, we will ignore the fact that some assets do not trade over the weekend.

Similarly, using results from chapter 2 about the additivity of covariances we have that covariance between ε_{it}^A and ε_{jt}^A is just 12 times the monthly covariance:

$$\begin{aligned}
cov(\varepsilon_{it}^A, \varepsilon_{jt}^A) &= cov\left(\sum_{k=0}^{11} \varepsilon_{it-k}, \sum_{k=0}^{11} \varepsilon_{jt-k}\right) \\
&= \sum_{k=0}^{11} cov(\varepsilon_{it-k}, \varepsilon_{jt-k}) \text{ since } \varepsilon_{it} \text{ and } \varepsilon_{jt} \text{ are uncorrelated over time} \\
&= \sum_{k=0}^{11} \sigma_{ij} \text{ since } cov(\varepsilon_{it}, \varepsilon_{jt}) \text{ is constant over time} \\
&= 12 \cdot \sigma_{ij} \\
&= cov(R_{it}^A, R_{jt}^A)
\end{aligned}$$

The above results imply that the correlation between ε_{it}^A and ε_{jt}^A is the same as the correlation between ε_{it} and ε_{jt} :

$$\begin{aligned}
corr(\varepsilon_{it}^A, \varepsilon_{jt}^A) &= \frac{cov(\varepsilon_{it}^A, \varepsilon_{jt}^A)}{\sqrt{var(\varepsilon_{it}^A) \cdot var(\varepsilon_{jt}^A)}} \\
&= \frac{12 \cdot \sigma_{ij}}{\sqrt{12\sigma_i^2 \cdot 12\sigma_j^2}} \\
&= \frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij} \\
&= corr(\varepsilon_{it}, \varepsilon_{jt})
\end{aligned}$$

1.0.4 The CER Model of Asset Returns and the Random Walk Model of Asset Prices

The CER model of asset returns (1.1) gives rise to the so-called *random walk* (RW) model of the *logarithm* of asset prices. To see this, recall that the continuously compounded return, R_{it}, is defined from asset prices via

$$\ln\left(\frac{P_{it}}{P_{it-1}}\right) = R_{it}.$$

Since the log of the ratio of prices is equal to the difference in the logs of prices we may rewrite the above as

$$\ln(P_{it}) - \ln(P_{it-1}) = R_{it}.$$

Letting $p_{it} = \ln(P_{it})$ and using the representation of R_{it} in the CER model (1.1), we may further rewrite the above as

$$p_{it} - p_{it-1} = \mu_i + \varepsilon_{it}. \tag{1.3}$$

The representation in (1.3) is know as the RW model for the log of asset prices.

In the RW model, μ_i represents the expected change in the log of asset prices (continuously compounded return) between months $t-1$ and t and ε_{it} represents the unexpected change in prices. That is,

$$\begin{aligned} E[p_{it} - p_{it-1}] &= E[R_{it}] = \mu_i, \\ \varepsilon_{it} &= p_{it} - p_{it-1} - E[p_{it} - p_{it-1}]. \end{aligned}$$

Further, in the RW model, the unexpected changes in asset prices, ε_{it}, are uncorrelated over time ($cov(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $t \neq s$) so that future changes in asset prices cannot be predicted from past changes in asset prices[3].

The RW model gives the following interpretation for the evolution of asset prices. Let p_{i0} denote the initial log price of asset i. The RW model says that the price at time $t = 1$ is

$$p_{i1} = p_{i0} + \mu_i + \varepsilon_{i1}$$

where ε_{i1} is the value of random news that arrives between times 0 and 1. Notice that at time $t = 0$ the expected price at time $t = 1$ is

$$E[p_{i1}] = p_{i0} + \mu_i + E[\varepsilon_{i1}] = p_{i0} + \mu_i$$

which is the initial price plus the expected return between time 0 and 1. Similarly, the price at time $t = 2$ is

$$\begin{aligned} p_{i2} &= p_{i1} + \mu_i + \varepsilon_{i2} \\ &= p_{i0} + \mu_i + \mu_i + \varepsilon_{i1} + \varepsilon_{i2} \\ &= p_{i0} + 2 \cdot \mu_i + \sum_{t=1}^{2} \varepsilon_{it} \end{aligned}$$

[3] The notion that future changes in asset prices cannot be predicted from past changes in asset prices is often referred to as the weak form of the efficient markets hypothesis.

which is equal to the initial price, p_{i0}, plus the two period expected return, $2 \cdot \mu_i$, plus the accumulated random news over the two periods, $\sum_{t=1}^{2} \varepsilon_{it}$. By recursive substitution, the price at time $t = T$ is

$$p_{iT} = p_{i0} + T \cdot \mu_i + \sum_{t=1}^{T} \varepsilon_{it}.$$

At time $t = 0$ the expected price at time $t = T$ is

$$E[p_{iT}] = p_{i0} + T \cdot \mu_i$$

The actual price, p_{iT}, deviates from the expected price by the accumulated random news

$$p_{iT} - E[p_{iT}] = \sum_{t=1}^{T} \varepsilon_{it}.$$

Figure 1.1 illustrates the random walk model of asset prices based on the CER model with $\mu = 0.05$, $\sigma = 0.10$ and $p_0 = 1$. The plot shows the log price, p_t, the expected price $E[p_t] = p_0 + 0.05t$ and the accumulated random news $\sum_{t=1}^{t} \varepsilon_t$.

The term *random walk* was originally used to describe the unpredictable movements of a drunken sailor staggering down the street. The sailor starts at an initial position, p_0, outside the bar. The sailor generally moves in the direction described by μ but randomly deviates from this direction after each step t by an amount equal to ε_t. After T steps the sailor ends up at position $p_T = p_0 + \mu \cdot T + \sum_{t=1}^{T} \varepsilon_t$.

1.1 Monte Carlo Simulation of the CER Model

A good way to understand the probabilistic behavior of a model is to use computer simulation methods to create pseudo data from the model. The process of creating such pseudo data is often called *Monte Carlo simulation*[4]. To illustrate the use of Monte Carlo simulation, consider the problem of creating pseudo return data from the CER model (1.1) for one asset. The steps to create a Monte Carlo simulation from the CER model are:

- Fix values for the CER model parameters μ and σ (or σ^2)

[4]Monte Carlo referrs to the fameous city in Monaco where gambling is legal.

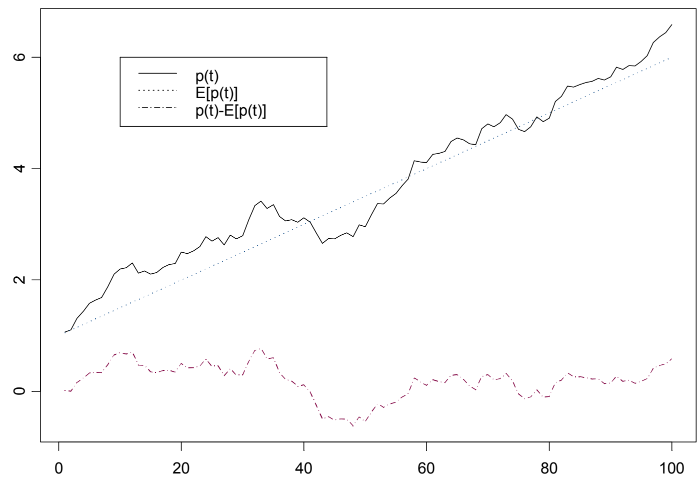


Figure 1.1: Simulated random walk model for log prices.

- Determine the number of simulated values, T, to create.
- Use a computer random number generator to simulate T *iid* values of ε_t from $N(0, \sigma^2)$ distribution. Denote these simulated values are $\varepsilon_1^*, \ldots, \varepsilon_T^*$.
- Create simulated return data $R_t^* = \mu + \varepsilon_t^*$ for $t = 1, \ldots, T$

To mimic the monthly return data on Microsoft, the values $\mu = 0.05$ and $\sigma = 0.10$ are used as the model's parameters and $T = 100$ is the number of simulated values (sample size). The key to simulating data from the above model is to simulate $T = 100$ observations of the random news variable ε_t $\tilde{}$*iid* $N(0, (0.10)^2)$. Computer algorithms exist which can easily create such observations..Let $\{\varepsilon_1^*, \ldots, \varepsilon_{100}^*\}$ denote the 100 simulated values of ε_t.The simulated returns are then computed as

$$R_t^* = 0.05 + \varepsilon_t^*, \ t = 1, \ldots, 100$$

A time plot and histogram of the simulated R_t^* values are given in figure .The simulated return data fluctuates randomly about the expected return

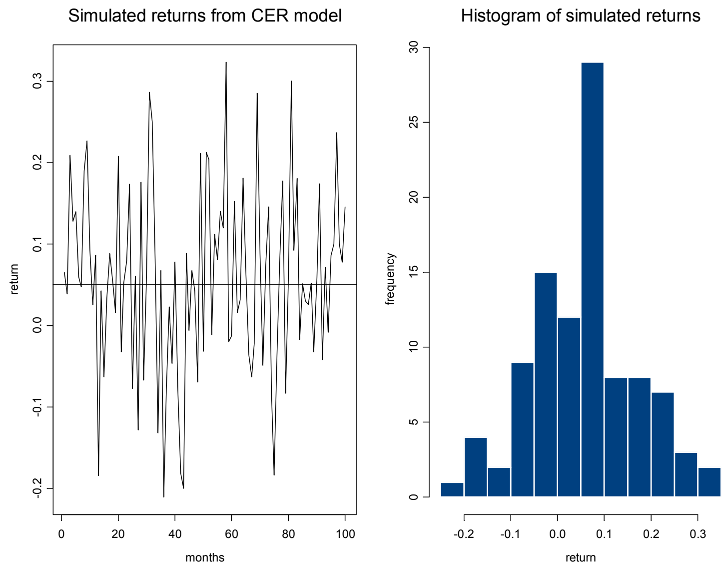


Figure 1.2: Simulated returns from the CER model $R_t = 0.05 + \varepsilon_t$, $\varepsilon_t \text{~} iid$ $N(0, (0.10)^2)$

value $E[R_t] = \mu = 0.05$. The typical size of the fluctuation is approximately equal to $SD(\varepsilon_t) = 0.10$. Notice that the simulated return data looks remarkably like the actual monthly return data for Microsoft.

The sample average of the simulated returns is $\frac{1}{100}\sum_{t=1}^{100} R_t^* = 0.0522$ and the sample standard deviation is $\sqrt{\frac{1}{99}\sum_{t=1}^{100}(R_t^* - (0.0522))^2} = 0.0914$. These values are very close to the population values $E[R_t] = 0.05$ and $SD(R_t) = 0.10$, respectively.

Monte Carlo simulation of a model can be used as a first pass reality check of the model. If simulated data from the model does not look like the data that the model is supposed to describe then serious doubt is cast on the model. However, if simulated data looks reasonably close to the data that the model is suppose to describe then confidence is instilled on the model.

1.1.1 Simulating End of Period Wealth

To be completed

- insert example showing how to use Monte Carlo simulation to compute expected end of period wealth. compare computations where end of period wealth is based on the expected return over the period versus computations based on simulating different sample paths and then taking the average. Essentially, compute $E[W_0 \exp(\sum_{t=1}^{N} R_t)]$ where R_t behaves according to the CER model and compare this to $W_0 \exp(N\mu)$.

1.1.2 Simulating Returns on More than One Asset

To be completed

1.2 Estimating the Parameters of the CER Model

1.2.1 The Random Sampling Environment

The CER model of asset returns gives us a rigorous way of interpreting the time series behavior of asset returns. At the beginning of every month t, R_{it} is a random variable representing the return to be realized at the end of the month. The CER model states that $R_{it} \sim i.i.d.\ N(\mu_i, \sigma_i^2)$. Our best guess for the return at the end of the month is $E[R_{it}] = \mu_i$, our measure of uncertainty about our best guess is captured by $\sigma_i = \sqrt{var(R_{it})}$ and our measure of the direction of linear association between R_{it} and R_{jt} is $\sigma_{ij} = cov(R_{it}, R_{jt})$. The CER model assumes that the economic environment is constant over time so that the normal distribution characterizing monthly returns is the same every month.

Our life would be very easy if we knew the exact values of μ_i, σ_i^2 and σ_{ij}, the parameters of the CER model. In actuality, however, we do not know these values with certainty. A key task in financial econometrics is estimating the values of μ_i, σ_i^2 and σ_{ij} from a history of observed data.

Suppose we observe monthly returns on N different assets over the horizon $t = 1, \ldots, T$. Let $\{r_{i1}, \ldots, r_{iT}\}$ denote the observed history of T monthly returns on asset i for $i = 1, \ldots, N$. It is assumed that the observed returns are realizations of the time series of random variables $\{R_{i1}, \ldots, R_{iT}\}$, where R_{it} is described by the CER model (1.1). We call $\{R_{i1}, \ldots, R_{iT}\}$ a *random sample* from the CER model (1.1) and we call $\{r_{i1}, \ldots, r_{iT}\}$ the *realized values* from the random sample. Under these assumptions, we can use the observed returns to estimate the unknown parameters of the CER model

1.2.2 Statistical Estimation Theory

Before we describe the estimation of the CER model, it is useful to summarize some concepts in the statistical theory of estimation. Let θ denote some characteristic of the CER model (1.1) we are interested in estimating. For example, if we are interested in the expected return then $\theta = \mu_i$; if we are interested in the variance of returns then $\theta = \sigma_i^2$. The goal is to estimate θ based on the observed data $\{r_{i1}, \ldots, r_{iT}\}$.

Definition 1 *An estimator of θ is a rule or algorithm for forming an estimate for θ based on the random sample $\{R_{i1}, \ldots, R_{iT}\}$*

Definition 2 *An estimate of θ is simply the value of an estimator based on the realized sample values $\{r_{i1}, \ldots, r_{iT}\}$.*

Example 3 *The sample average $\frac{1}{T}\sum_{t=1}^{T} R_{it}$ is an algorithm for computing an estimate of the expected return μ_i. Before the sample is observed, the sample average is a simple linear function of the random variables $\{R_{i1}, \ldots, R_{iT}\}$ and so is itself a random variable. After the sample $\{r_{i1}, \ldots, r_{iT}\}$ is observed, the sample average can be evaluated giving $\frac{1}{T}\sum_{t=1}^{T} r_{it}$, which is just a number. For example, if the observed sample is $\{0.05, 0.03, -0.10\}$ then the sample average estimate is $\frac{1}{3}(0.05 + 0.03 - 0.10) = -0.02$.*

To discuss the properties of estimators it is necessary to establish some notation. Let $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ denote an estimator of θ treated as a function of the random variables $\{R_{i1}, \ldots, R_{iT}\}$. Clearly, $\hat{\theta}(R_{i1}, \ldots, R_{iT})$ is a random variable. Let $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ denote an estimate of θ based on the realized values $\{r_{i1}, \ldots, r_{iT}\}$. $\hat{\theta}(r_{i1}, \ldots, r_{iT})$ is simply an number. We will often use $\hat{\theta}$ as shorthand notation to represent either an estimator of θ or an estimate of θ. The context will determine how to interpret $\hat{\theta}$.

Example 4 *Let $R_1, \ldots, R_T$ denote a random sample of returns. An estimator of the expected return, μ, is the sample average*

$$\hat{\mu}(R_1, \ldots, R_T) = \frac{1}{T}\sum_{t=1}^{T} R_t$$

Suppose $T = 5$ *and the realized values of the returns are* $r_1 = 0.1, r_2 = 0.05, r_3 = 0.025, r_4 = -0.1, r_5 = -0.05$. *Then the estimate of the expected return using the sample average is*

$$\hat{\mu}(0.1, \ldots, -0.05) = \frac{1}{5}(0.1 + 0.05 + 0.025 + -0.1 + -0.05) = 0.005$$

1.2.3 Properties of Estimators

Consider $\hat{\theta} = \hat{\theta}(R_{i1}, \ldots, R_{iT})$ as a random variable. In general, the pdf of $\hat{\theta}$, $p(\hat{\theta})$, depends on the pdf's of the random variables $R_{i1}, \ldots, R_{iT}$. The exact form of $p(\hat{\theta})$ may be very complicated. For analysis purposes, we often focus on certain characteristics of $p(\hat{\theta})$ like its expected value (center), variance and standard deviation (spread about expected value). The expected value of an estimator is related to the concept of estimator *bias* and the variance/standard deviation of an estimator is related estimator *precision.* Intuitively, a *good* estimator of θ is one that will produce an estimate $\hat{\theta}$ that is close θ all of the time. That is, a good estimator will have small bias and high precision.

Bias

Bias concerns the location or center of $p(\hat{\theta})$. If $p(\hat{\theta})$ is centered away from θ then we say $\hat{\theta}$ is *biased.* If $p(\hat{\theta})$ is centered at θ then we say that $\hat{\theta}$ is *unbiased.* Formally we have the following definitions:

Definition 5 *The estimation error is difference between the estimator and the parameter being estimated*

$$error = \hat{\theta} - \theta.$$

Definition 6 *The bias of an estimator* $\hat{\theta}$ *of* θ *is given by*

$$bias(\hat{\theta}, \theta) = E[\hat{\theta}] - \theta.$$

Definition 7 *An estimator* $\hat{\theta}$ *of* θ *is unbiased if* $bias(\hat{\theta}, \theta) = 0$; *i.e., if* $E[\hat{\theta}] = \theta$ *or* $E[error] = 0$.

Unbiasedness is a desirable property of an estimator. It means that the estimator produces the correct answer "on average", where "on average"

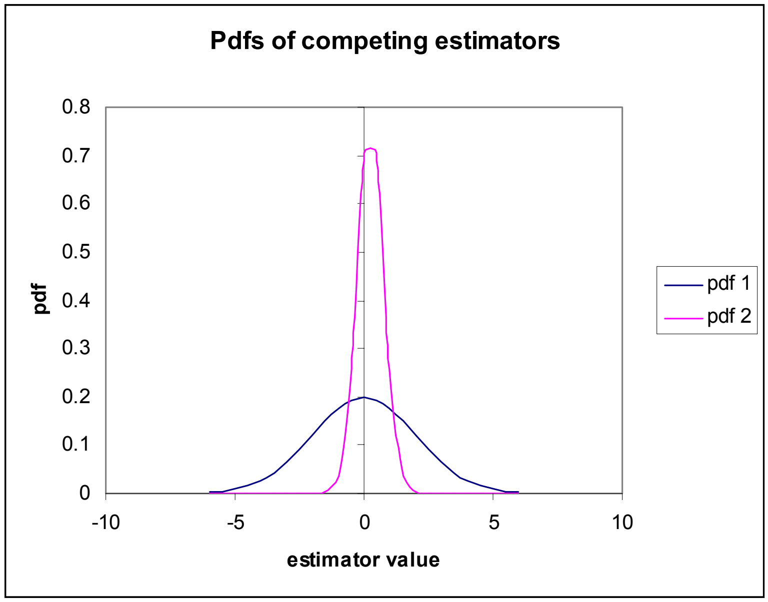


Figure 1.3: Pdf values for competing estimators of $\theta = 0$.

means over many hypothetical samples. It is important to keep in mind that an unbiased estimator for θ may not be very close to θ for a particular sample and that a biased estimator may be actually be quite close to θ. For example, consider the pdf of $\hat{\theta}_1$ in figure 1.3. The center of the distribution is at the true value $\theta = 0, E[\hat{\theta}_1] = 0$, but the distribution is very widely spread out about $\theta = 0$. That is, $var(\hat{\theta}_1)$ is large. On average (over many hypothetical samples) the value of $\hat{\theta}_1$ will be close to θ but in any given sample the value of $\hat{\theta}_1$ can be quite a bit above or below θ. Hence, unbiasedness by itself does not guarantee a good estimator of θ. Now consider the pdf for $\hat{\theta}_2$. The center of the pdf is slightly higher than $\theta = 0, bias(\hat{\theta}_2, \theta) = 0.25$, but the spread of the distribution is small. Although the value of $\hat{\theta}_2$ is not equal to 0 *on average* we might prefer the estimator $\hat{\theta}_2$ over $\hat{\theta}_1$ because it is generally closer to $\theta = 0$ *on average* than $\hat{\theta}_1$.

Precision

An estimate is, hopefully, our best guess of the true (but unknown) value of θ. Our guess most certainly will be wrong but we hope it will not be too far off. A precise estimate, loosely speaking, is one that has a small estimation error. The magnitude of the estimation error is usually captured by the *mean squared error:*

Definition 8 *The mean squared error of an estimator $\hat{\theta}$ of θ is given by*

$$mse(\hat{\theta}, \theta) = E[(\hat{\theta} - \theta)^2] = E[error^2]$$

The mean squared error measures the expected squared deviation of $\hat{\theta}$ from θ. If this expected deviation is small, then we know that $\hat{\theta}$ will almost always be close to θ. Alternatively, if the mean squared is large then it is possible to see samples for which $\hat{\theta}$ to be quite far from θ. A useful decomposition of $mse(\hat{\theta}, \theta)$ is given in the following proposition

Proposition 9 $mse(\hat{\theta}, \theta) = E[(\hat{\theta} - E[\hat{\theta}])^2] + \left(E[\hat{\theta}] - \theta\right)^2 = var(\hat{\theta}) + bias(\hat{\theta}, \theta)^2$

The proof of this proposition is straightforward and is given in the appendix. The proposition states that for any estimator $\hat{\theta}$ of θ, $mse(\hat{\theta}, \theta)$ can be split into a variance component, $var(\hat{\theta})$, and a bias component, $bias(\hat{\theta}, \theta)^2$. Clearly, $mse(\hat{\theta}, \theta)$ will be small only if both components are small. If an estimator is unbiased then $mse(\hat{\theta}, \theta) = var(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$ is just the squared deviation of $\hat{\theta}$ about θ. Hence, an unbiased estimator $\hat{\theta}$ of θ is *good* if it has a small variance.

1.2.4 Method of Moment Estimators for the Parameters of the CER Model

Let $\{R_{i1}, \ldots, R_{iT}\}$ denote a random sample from the CER model and let $\{r_{i1}, \ldots, r_{iT}\}$ denote the realized values from the random sample. Consider the problem of estimating the parameter μ_i in the CER model (1.1). As an example, consider the observed monthly continuously compounded returns, $\{r_1, \ldots, r_{100}\}$, for Microsoft stock over the period July 1992 through October 2000. These data are illustrated in figure 1.4.Notice that the data seem to fluctuate up and down about some central value near 0.03. The typical size of a deviation about 0.03 is roughly 0.10. Intuitively, the parameter $\mu_i = E[R_{it}]$ in the CER model represents this central value and σ_i represents the typical size of a deviation about μ_i.

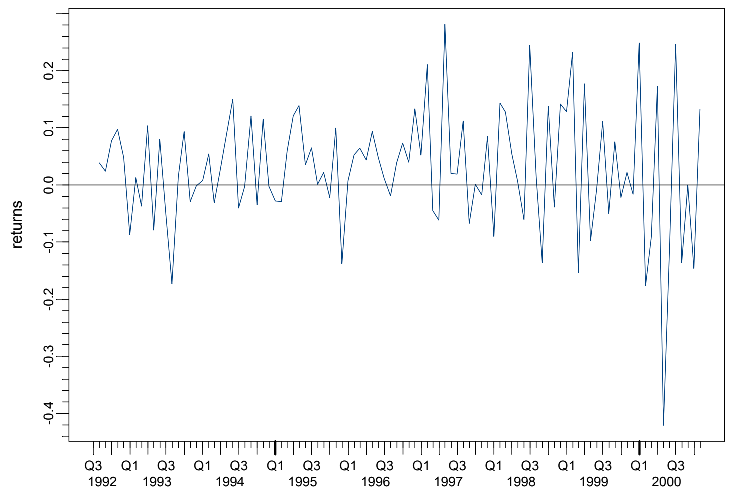


Figure 1.4: Monthly continuously compounded returns on Microsoft stock.

The method of moments estimate of μ_i

Let $\hat{\mu}_i$ denote a prospective estimate of μ_i[5]. The *sample error* or *residual* at time t associated with this estimate is defined as

$$\hat{\varepsilon}_{it} = r_{it} - \hat{\mu}_i, \;\; t = 1, \ldots, T.$$

This is the estimated news component for month t based on the estimate $\hat{\mu}_i$. Now the CER model imposes the condition that the expected value of the true error is zero

$$E[\varepsilon_{it}] = 0$$

The *method of moments* estimator of μ_i is the value of $\hat{\mu}_i$ that makes the average of the sample errors equal to the expected value of the population errors. That is, the method of moments estimator solves

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it} = \frac{1}{T}\sum_{t=1}^{T}(r_{it} - \hat{\mu}_i) = E[\varepsilon_{it}] = 0 \tag{1.4}$$

[5] In this book, quantities with a "ˆ" denote an estimate.

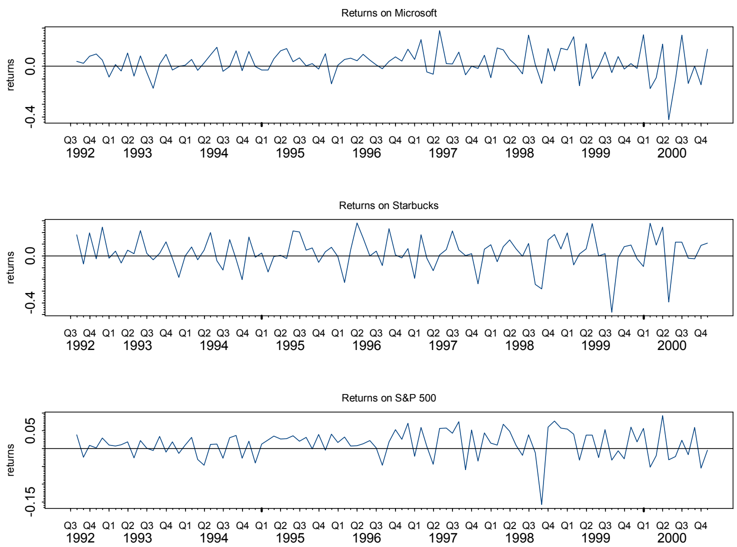


Figure 1.5: Monthly continuously compounded returns on Microsoft, Starbucks and the S&P 500 Index.

Solving (1.4) for $\hat{\mu}_i$ gives the method of moments estimate of μ_i :

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it} = \bar{r}. \tag{1.5}$$

Hence, the method of moments estimate of μ_i $(i = 1, \ldots, N)$ in the CER model is simply the *sample average* of the observed returns for asset i.

Example 10 *Consider the monthly continuously compounded returns on Microsoft, Starbucks and the S&P 500 index over the period July 1992 through October 2000. The returns are shown in figure For the* $T = 100$ *monthly* *continuously returns the estimates of* $E[R_{it}] = \mu_i$ *are*

$$\begin{aligned}\hat{\mu}_{msft} &= \frac{1}{100}\sum_{t=1}^{100} r_{msft,t} = 0.0276\\ \hat{\mu}_{sbux} &= \frac{1}{100}\sum_{t=1}^{100} r_{sbux,t} = 0.0278\\ \hat{\mu}_{sp500} &= \frac{1}{100}\sum_{t=1}^{100} r_{sp500,t} = 0.0125\end{aligned}$$

The mean returns for MSFT and SBUX are very similar at about 2.8% per month whereas the mean return for SP500 is smaller at only 1.25% per month.

The method of moments estimates of σ_i^2, σ_i, σ_{ij} and ρ_{ij}

The method of moments estimates of σ_i^2, σ_i, σ_{ij} and ρ_{ij} are defined analogously to the method of moments estimator for μ_i. Without going into the details, the method of moments estimates of σ_i^2, σ_i, σ_{ij} and ρ_{ij} are given by the sample descriptive statistics

$$\hat{\sigma}_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it} - \bar{r}_i)^2, \tag{1.6}$$

$$\hat{\sigma}_i = \sqrt{\hat{\sigma}_i^2}, \tag{1.7}$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1}\sum_{t=1}^{T}(r_{it} - \bar{r}_i)(r_{jt} - \bar{r}_j), \tag{1.8}$$

$$\hat{\rho}_{ij} = \frac{\hat{\sigma}_{ij}}{\hat{\sigma}_i\hat{\sigma}_j} \tag{1.9}$$

where $\bar{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} = \hat{\mu}_i$ is the sample average of the returns on asset.i. Notice that (1.6) is simply the *sample variance* of the observed returns for asset i, (1.7) is the sample standard deviation, (1.8) is the *sample covariance* of the observed returns on assets i and j and (1.9) is the sample correlation of returns on assets i and j.

Example 11 *Consider again the monthly continuously compounded returns on Microsoft, Starbucks and the S&P 500 index over the period July 1992*

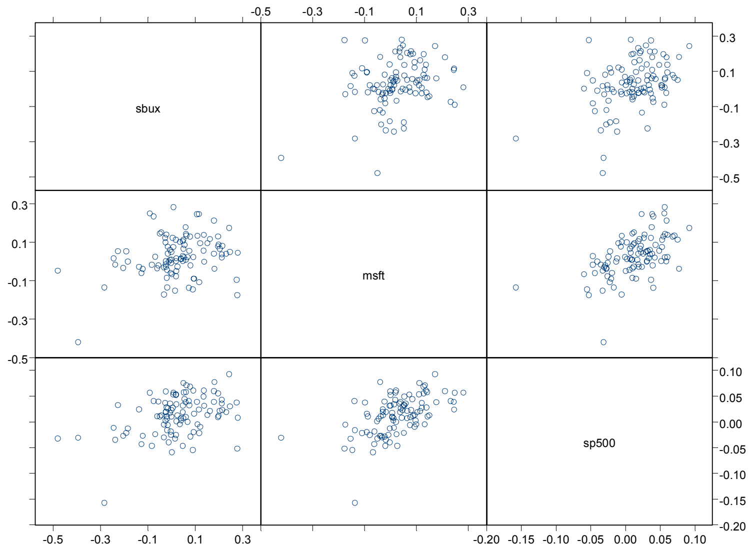


Figure 1.6: Scatterplot matrix of monthly returns on Microsoft, Starbucks and S&P 500 index.

through October 2000. The estimates of the parameters σ_i^2, σ_i, using (1.6) and (1.7) are

$$
\begin{aligned}
\hat{\sigma}^2_{msft} &= 0.0114, \hat{\sigma}_{msft} = 0.1068 \\
\hat{\sigma}^2_{sbux} &= 0.0185, \hat{\sigma}_{sbux} = 0.1359 \\
\hat{\sigma}^2_{sp500} &= 0.0014, \hat{\sigma}_{sp500} = 0.0379
\end{aligned}
$$

SBUX has the most variable monthly returns and SP500 has the smallest. The scatterplots of the returns are illustrated in figure 1.6. All returns appear to be positively related. The pairs (MSFT,SP500) and (SBUX,SP500) appear to be the most correlated.The estimates of σ_{ij} and ρ_{ij} using (1.8) and (1.9) are

$$
\begin{aligned}
\hat{\sigma}_{msft,sbux} &= 0.0040, \hat{\sigma}_{msft,sp500} = 0.0022, \hat{\sigma}_{sbux,sp500} = 0.0022 \\
\hat{\rho}_{msft,sbux} &= 0.2777, \hat{\rho}_{msft,sp500} = 0.5551, \hat{\rho}_{sbux,sp500} = 0.4197
\end{aligned}
$$

These estimates confirm the visual results from the scatterplot matrix.

1.3 Statistical Properties of Estimates

1.3.1 Statistical Properties of $\hat{\mu}_i$

To determine the statistical properties of $\hat{\mu}_i$ in the CER model, we treat it as a function of the random sample $R_{i1}, \ldots, R_{iT}$:

$$\hat{\mu}_i = \hat{\mu}_i(R_{i1}, \ldots, R_{iT}) = \frac{1}{T}\sum_{t=1}^{T} R_{it} \tag{1.10}$$

where R_{it} is assumed to be generated by the CER model (1.1).

Bias

In the CER model, the random variables R_{it} $(t = 1, \ldots, T)$ are *iid* normal with mean μ_i and variance σ_i^2. Since the method of moments estimator (1.10) is an average of these normal random variables it is also normally distributed. That is, $p(\hat{\mu}_i)$ is a normal density. To determine the mean of this distribution we must compute $E[\hat{\mu}_i] = E[T^{-1}\sum_{t=1}^{T} R_{it}]$. Using results from chapter 2 about the expectation of a linear combination of random variables it is straightforward to show (details are given in the appendix) that

$$E[\hat{\mu}_i] = \mu_i$$

Hence, the mean of the distribution of $\hat{\mu}_i$ is equal to μ_i. In other words, $\hat{\mu}_i$ an *unbiased estimator* for μ_i.

Precision

To determine the variance of $\hat{\mu}_i$ we must compute $var(\hat{\mu}_i) = var(T^{-1}\sum_{t=1}^{T} R_{it})$. Using the results from chapter 2 about the variance of a linear combination of uncorrelated random variables it is easy to show (details in the appendix) that

$$var(\hat{\mu}_i) = \frac{\sigma^2}{T}. \tag{1.11}$$

Notice that the variance of $\hat{\mu}_i$ is equal to the variance of R_{it} divided by the sample size and is therefore much smaller than the variance of R_{it}.

The standard deviation of $\hat{\mu}_i$ is just the square root of $var(\hat{\mu}_{it})$

$$SD(\hat{\mu}_i) = \sqrt{var(\hat{\mu}_i)} = \frac{\sigma_i}{\sqrt{T}}. \tag{1.12}$$

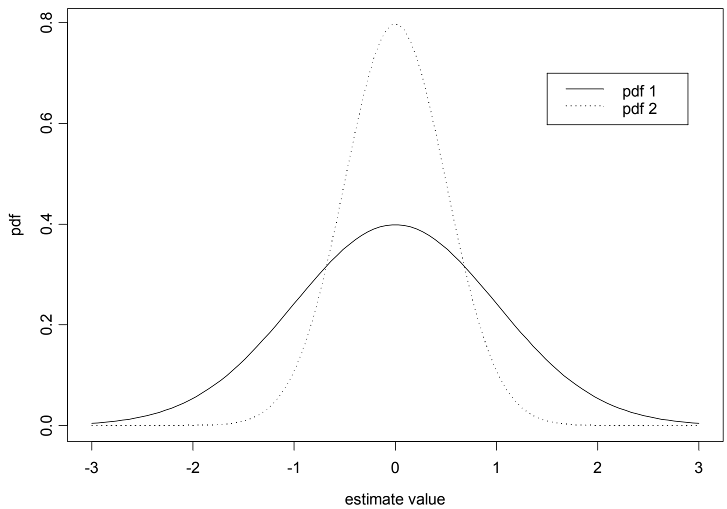


Figure 1.7: Pdfs for $\hat{\mu}_i$ with small and large values of $SE(\hat{\mu}_i)$. True value of $\mu_i = 0$.

The standard deviation of $\hat{\mu}_i$ is most often referred to as the *standard error* of the estimate $\hat{\mu}_i$:

$$SE(\hat{\mu}_i) = SD(\hat{\mu}_i) = \frac{\sigma_i}{\sqrt{T}}. \tag{1.13}$$

$SE(\hat{\mu}_i)$ is in the same units as $\hat{\mu}_i$ and measures the precision of $\hat{\mu}_i$ as an estimate. If $SE(\hat{\mu}_i)$ is small relative to $\hat{\mu}_i$ then $\hat{\mu}_i$ is a relatively precise of μ_i because $p(\hat{\mu}_i)$ will be tightly concentrated around μ_i; if $SE(\hat{\mu}_i)$ is large relative to μ_i then $\hat{\mu}_i$ is a relatively imprecise estimate of μ_i because $p(\hat{\mu}_i)$ will be spread out about μ_i. Figure 1.7 illustrates these relationships

Unfortunately, $SE(\hat{\mu}_i)$ is not a *practically useful* measure of the precision of $\hat{\mu}_i$ because it depends on the unknown value of σ_i. To get a practically useful measure of precision for $\hat{\mu}_i$ we compute the *estimated standard error*

$$\widehat{SE}(\hat{\mu}_i) = \sqrt{\widehat{var}(\hat{\mu}_i)} = \frac{\widehat{\sigma}_i}{\sqrt{T}} \tag{1.14}$$

which is just (1.13) with the unknown value of σ_i replaced by the method of moments estimate $\widehat{\sigma}_i = \sqrt{\widehat{\sigma}_i^2}$.

Example 12 *For the Microsoft, Starbucks and S&P 500 return data, the values of $\widehat{SE}(\hat{\mu}_i)$ are*

$$\begin{aligned}
\widehat{SE}(\hat{\mu}_{msft}) &= \frac{0.1068}{\sqrt{100}} = 0.01068 \\
\widehat{SE}(\hat{\mu}_{sbux}) &= \frac{0.1359}{\sqrt{100}} = 0.01359 \\
\widehat{SE}(\hat{\mu}_{sp500}) &= \frac{0.0379}{\sqrt{100}} = 0.003785
\end{aligned}$$

Clearly, the mean return μ_i is estimated more precisely for the S&P 500 index than it is for Microsoft and Starbucks.

Interpreting $E[\hat{\mu}_i]$ and $SE(\hat{\mu}_i)$ using Monte Carlo simulation

The statistical concepts $E[\mu_i] = \mu_i$ and $SE(\mu_i)$ are a bit hard to grasp at first. Strictly speaking, $E[\hat{\mu}_i] = \mu_i$ means that over an infinite number of repeated samples the average of the $\hat{\mu}_i$ values computed over the infinite samples is equal to the true value μ_i. Similarly, $SE(\hat{\mu}_i)$ represents the standard deviation of these $\hat{\mu}_i$ values. We may think of these hypothetical samples as Monte Carlo simulations of the CER model. In this way we can approximate the computations involved in evaluating $E[\hat{\mu}_i]$ and $SE(\hat{\mu}_i)$.

To illustrate, consider the CER model

$$\begin{aligned}
R_t &= 0.05 + \varepsilon_{it}, t = 1, \ldots, 50 \\
\varepsilon_{it} &\,\tilde{}\, iid\ N(0, (0.10)^2)
\end{aligned} \tag{1.15}$$

and simulate $N = 1000$ samples of size $T = 50$ values from the above model using the technique of Monte Carlo simulation. This gives $j = 1, \ldots, 1000$ sample realizations $\{r_1^{j*}, \ldots, r_{50}^{j*}\}$. The first 10 of these sample realizations are illustrated in figure 1.8.Notice that there is considerable variation in the simulated samples but that all of the simulated samples fluctuates about the true mean value of $\mu = 0.05$. For each of the 1000 simulated samples the estimate $\hat{\mu}$ is formed giving 1000 mean estimates $\{\hat{\mu}^1, \ldots, \hat{\mu}^{1000}\}$. A histogram of these 1000 mean values is illustrated in figure 1.9.The histogram of the estimated means, $\hat{\mu}^j$, can be thought of as an estimate of the underlying pdf, $p(\hat{\mu})$, of the estimator $\hat{\mu}$ which we know is a Normal pdf centered at $\mu = 0.05$ with $SE(\hat{\mu}_i) = \frac{0.10}{\sqrt{50}} = 0.01414$. Notice that the center of the histogram is very close to the true mean value $\mu = 0.05$. That is, on average over the

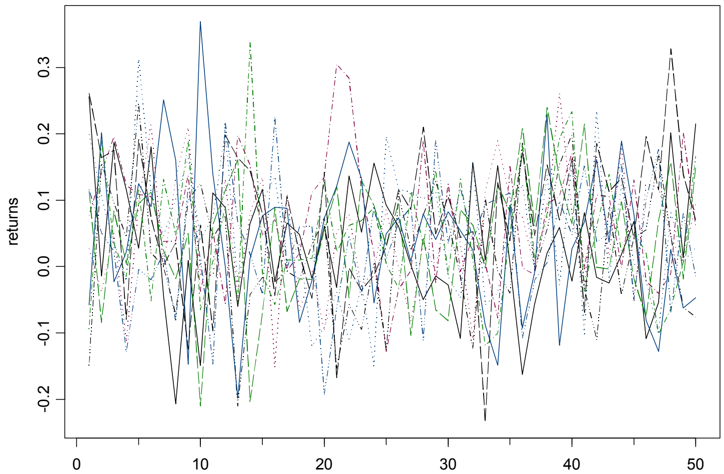


Figure 1.8: Ten simulated samples of size $T = 50$ from the CER model $R_t = 0.05 + \varepsilon_t, \varepsilon_t \,\tilde{}\, iid\ N(0.(0.10)^2)$

1000 Monte Carlo samples the value of $\hat{\mu}$ is about 0.05. In some samples, the estimate is too big and in some samples the estimate is too small but *on average* the estimate is correct. In fact, the average value of $\{\hat{\mu}^1, \ldots, \hat{\mu}^{1000}\}$ from the 1000 simulated samples is

$$\frac{1}{1000}\sum_{j=1}^{1000}\hat{\mu}^j = 0.05045$$

which is very close to the true value. If the number of simulated samples is allowed to go to infinity then the sample average of $\hat{\mu}^j$ will be exactly equal to μ :

$$\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}\hat{\mu}^j = \mu$$

The typical size of the spread about the center of the histogram represents $SE(\hat{\mu}_i)$ and gives an indication of the precision of $\hat{\mu}_i$.The value of $SE(\hat{\mu}_i)$ may be approximated by computing the sample standard deviation of the 1000

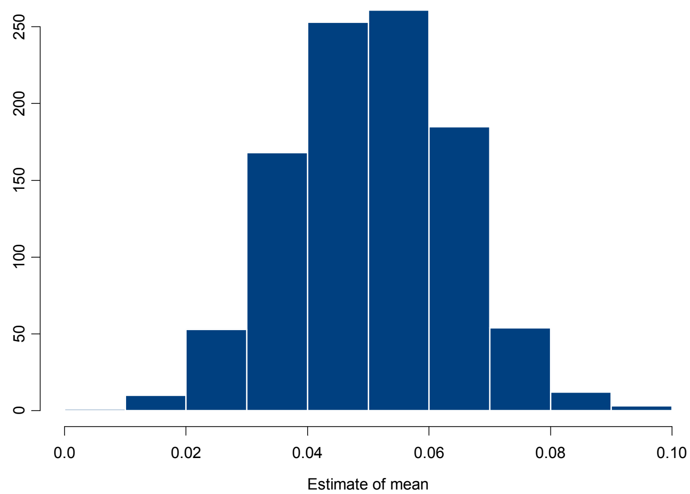


Figure 1.9: Histogram of 1000 values of $\hat{\mu}$ from Monte Carlo simulation of CER model.

$\hat{\mu}^j$ values

$$\sqrt{\frac{1}{999}\sum_{j=1}^{1000}(\hat{\mu}^j - 0.05045)^2} = 0.01383$$

Notice that this value is very close to $SE(\hat{\mu}_i) = \frac{0.10}{\sqrt{50}} = 0.01414$. If the number of simulated sample goes to infinity then

$$\lim_{N\to\infty}\sqrt{\frac{1}{N-1}\sum_{j=1}^{N}(\hat{\mu}^j - \frac{1}{N}\sum_{j=1}^{N}\hat{\mu}^j)^2} = SE(\hat{\mu}_i)$$

The Sampling Distribution of $\hat{\mu}_i$

Using the results that pdf of $\hat{\mu}_i$ is normal with $E[\hat{\mu}_i] = \mu_i$ and $var(\hat{\mu}_i) = \frac{\sigma_i^2}{T}$ we may write

$$\hat{\mu}_i \sim N\left(\mu_i, \frac{\sigma_i^2}{T}\right). \tag{1.16}$$

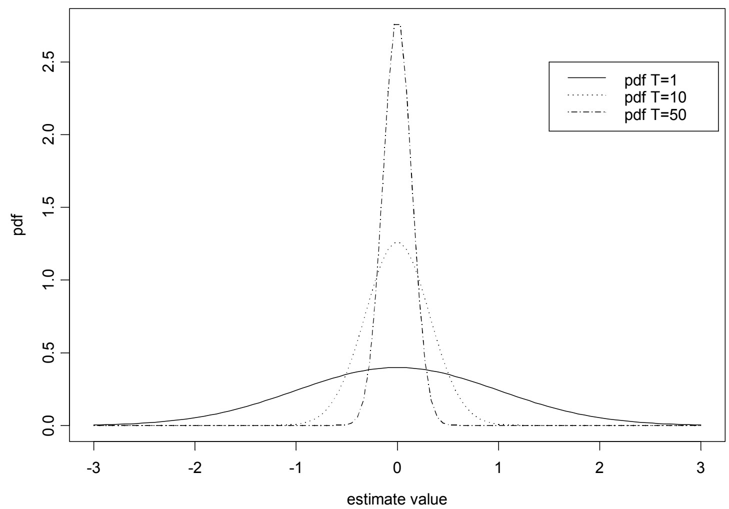


Figure 1.10: $N(0, \frac{1}{\sqrt{T}})$ density for $T = 1, 10$ and 50.

The distribution for $\hat{\mu}_i$ is centered at the true value μ_i and the spread about the average depends on the magnitude of σ_i^2, the variability of R_{it}, and the sample size. For a fixed sample size, T, the uncertainty in $\hat{\mu}_i$ is larger for larger values of σ_i^2. Notice that the variance of $\hat{\mu}_i$ is inversely related to the sample size T. Given σ_i^2, $var(\hat{\mu}_i)$ is smaller for larger sample sizes than for smaller sample sizes. This makes sense since we expect to have a more precise estimator when we have more data. If the sample size is very large (as $T \to \infty$) then $var(\hat{\mu}_i)$ will be approximately zero and the normal distribution of $\hat{\mu}_i$ given by (1.16) will be essentially a spike at μ_i. In other words, if the sample size is very large then we essentially know the true value of μ_i. In the statistics language we say that $\hat{\mu}_i$ is a *consistent* estimator of μ_i.

The distribution of $\hat{\mu}_i$, with $\mu_i = 0$ and $\sigma_i^2 = 1$ for various sample sizes is illustrated in figure 1.10. Notice how fast the distribution collapses at $\mu_i = 0$ as T increases. .

Confidence intervals for μ_i

The precision of $\hat{\mu}_i$ is best communicated by computing a *confidence interval* for the unknown value of μ_i. A confidence interval is an interval estimate of μ_i such that we can put an explicit probability statement about the likelihood that the confidence interval covers μ_i. The construction of a confidence interval for μ_i is based on the following statistical result (see the appendix for details).

Result: Let $R_{i1}, \ldots, R_{iT}$ denote a random sample from the CER model. Then

$$\frac{\hat{\mu}_i - \mu_i}{\widehat{SE}(\hat{\mu}_i)} \sim t_{T-1},$$

where t_{T-1} denotes a Student-t random variable with $T-1$ degrees of freedom.

The above result states that the standardized value of $\hat{\mu}_i$ has a Student-t distribution with $T-1$ degrees of freedom[6]. To compute a $(1-\alpha)\cdot 100\%$ confidence interval for μ_i we use (**??**) and the quantile (critical value) $t_{T-1}(\alpha/2)$ to give

$$\Pr\left(-t_{T-1}(\alpha/2) \leq \frac{\hat{\mu}_i - \mu_i}{\widehat{SE}(\hat{\mu}_i)} \leq t_{T-1}(\alpha/2)\right) = 1-\alpha,$$

which can be rearranged as

$$\Pr\left(\hat{\mu}_i - t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i) \leq \mu_i \leq \hat{\mu}_i + t_{T-1}\cdot\widehat{SE}(\hat{\mu}_i)\right) = 0.95.$$

Hence, the interval

$$[\hat{\mu}_i - t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i),\ \hat{\mu}_i + t_{T-1}\cdot\widehat{SE}(\hat{\mu}_i)] = \hat{\mu}_i \pm t_{T-1}(\alpha/2)\cdot\widehat{SE}(\hat{\mu}_i)$$

covers the true unknown value of μ_i with probability $1-\alpha$.

For example, suppose we want to compute 95% confidence intervals for μ_i. In this case $\alpha = 0.05$ and $1-\alpha = 0.95$. Suppose further that $T-1 = 60$ (five years of monthly return data) so that $t_{T-1}(\alpha/2) = t_{60}(0.025) = 2$ and $t_{60}(0.005) =$. Then the 95% confidence for μ_i is given by

$$\hat{\mu}_i \pm 2\cdot\widehat{SE}(\hat{\mu}_i). \tag{1.17}$$

[6] This resut follows from the fact that $\hat{\mu}_i$ is normally distributed and $\widehat{SE}(\hat{\mu}_i)$ is equal to the square root of a chi-square random variable divided by its degrees of freedom.

The above formula for a 95% confidence interval is often used as a rule of thumb for computing an approximate 95% confidence interval for moderate sample sizes. It is easy to remember and does not require the computation of quantile $t_{T-1}(\alpha/2)$ from the Student-t distribution.

Example 13 *Consider computing approximate 95% confidence intervals for* μ_i *using (1.17) based on the estimated results for the Microsoft, Starbucks and S&P 500 data. These confidence intervals are*

$$\begin{aligned} MSFT &: 0.02756 \pm 2 \cdot 0.01068 = [0.0062, 0.0489] \\ SBUX &: 0.02777 \pm 2 \cdot 0.01359 = [0.0006, 0.0549] \\ SP500 &: 0.01253 \pm 2 \cdot 0.003785 = [0.0050, 0.0201] \end{aligned}$$

With probability .95, the above intervals will contain the true mean values assuming the CER model is valid. The approximate 95% confidence intervals for MSFT and SBUX are fairly wide. The widths are almost 5% with lower limits near 0 and upper limits near 5%. In contrast, the 95% confidence interval for SP500 is about half the width of the MSFT or SBUX confidence interval. The lower limit is near .5% and the upper limit is near 2%. This clearly shows that the mean return for SP500 is estimated much more precisely than the mean return for MSFT or SBUX.

1.3.2 Statistical properties of the method of moments estimators of σ_i^2, σ_i, σ_{ij} and ρ_{ij}.

To determine the statistical properties of $\hat{\sigma}_i^2$ and $\hat{\sigma}_i^2$ we need to treat them as a functions of the random sample $R_{i1}, \ldots, R_{iT}$:

$$\begin{aligned} \hat{\sigma}_i^2 &= \hat{\sigma}_i^2(R_{i1}, \ldots R_{iT}) = \frac{1}{T-1}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2, \\ \hat{\sigma}_i &= \hat{\sigma}_i(R_{i1}, \ldots R_{iT}) = \sqrt{\hat{\sigma}_i^2(R_{i1}, \ldots R_{iT})}. \end{aligned}$$

Note also that $\hat{\mu}_i$ is to be treated as a random variable. Similarly, to determine the statistical properties of $\hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ we need to treat them as a functions of $R_{i1}, \ldots, R_{iT}$ and $R_{i1}, \ldots, R_{jT}$:

$$
\begin{aligned}
\hat{\sigma}_{ij} &= \hat{\sigma}_{ij}(R_{i1}, \ldots, R_{iT}; R_{j1}, \ldots, R_{jT}) = \frac{1}{T-1}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)(R_{jt} - \hat{\mu}_j), \\
\hat{\rho}_{ij} &= \hat{\sigma}_{ij}(R_{i1}, \ldots, R_{iT}; R_{j1}, \ldots, R_{jT}) = \frac{\hat{\sigma}_{ij}(R_{i1}, \ldots, R_{iT}; R_{j1}, \ldots, R_{jT})}{\hat{\sigma}_i(R_{i1}, \ldots R_{iT}) \cdot \hat{\sigma}_j(R_{j1}, \ldots R_{jT})}.
\end{aligned}
$$

Bias

Assuming that returns are generated by the CER model (1.1), the sample variances and covariances are unbiased estimators,

$$
\begin{aligned}
E[\hat{\sigma}_i^2] &= \sigma_i^2, \\
E[\hat{\sigma}_{ij}] &= \sigma_{ij},
\end{aligned}
$$

but the sample standard deviations and correlations are biased estimators,

$$
\begin{aligned}
E[\hat{\sigma}_i] &\neq \sigma_i, \\
E[\hat{\rho}_{ij}] &\neq \rho_{ij}.
\end{aligned}
$$

The proofs of these results are beyond the scope of this book. However, they may be easily be evaluated using Monte Carlo methods.

Precision

The derivations of the variances of $\hat{\sigma}_i^2, \hat{\sigma}_i$, $\hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ are complicated and the exact results are extremely messy and hard to work with. However, there are simple approximate formulas for the variances of $\hat{\sigma}_i^2, \hat{\sigma}_i$ and $\hat{\rho}_{ij}$ that are valid if the sample size, T, is reasonably large [7]. These large sample approximate formulas are given by

$$
\begin{aligned}
SE(\hat{\sigma}_i^2) &\approx \frac{\sigma_i^2}{\sqrt{T/2}}, && (1.18) \\
SE(\hat{\sigma}_i) &\approx \frac{\sigma_i}{\sqrt{2T}}, && (1.19) \\
SE(\rho_{ij}) &\approx \frac{(1-\rho_{ij}^2)}{\sqrt{T}}, && (1.20)
\end{aligned}
$$

[7] The large sample approximate formula for the variance of $\hat{\sigma}_{ij}$ is too messy to work with so we omit it here.

where "$\approx$" denotes approximately equal. The approximations are such that the approximation error goes to zero as the sample size T gets very large. As with the formula for the standard error of the sample mean, the formulas for the standard errors above are inversely related to the square root of the sample size. Interestingly, $SE(\hat{\sigma}_i)$ goes to zero the fastest and $SE(\hat{\sigma}_i^2)$ goes to zero the slowest. Hence, for a fixed sample size, it appears that σ_i is generally estimated more precisely than σ_i^2 and ρ_{ij}, and ρ_{ij} is estimated generally more precisely than σ_i^2.

The above formulas are not practically useful, however, because they depend on the unknown quantities σ_i^2, σ_i and ρ_{ij}. Practically useful formulas replace σ_i^2, σ_i and ρ_{ij} by the estimates $\hat{\sigma}_i^2, \hat{\sigma}_i$ and $\hat{\rho}_{ij}$ and give rise to the estimated standard errors

$$\widehat{SE}(\hat{\sigma}_i^2) \approx \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \tag{1.21}$$

$$\widehat{SE}(\hat{\sigma}_i) \approx \frac{\hat{\sigma}_i}{\sqrt{2T}}, \tag{1.22}$$

$$\widehat{SE}(\rho_{ij}) \approx \frac{(1-\hat{\rho}_{ij}^2)}{\sqrt{T}}. \tag{1.23}$$

Example 14 *To be completed*

Sampling distribution

To be completed

Confidence Intervals for σ_i^2, σ_i and ρ_{ij}

Approximate 95% confidence intervals for σ_i^2, σ_i and ρ_{ij} are give by

$$\begin{aligned}
\hat{\sigma}_i^2 \pm 2 \cdot \widehat{SE}(\hat{\sigma}_i^2) &= \hat{\sigma}_i^2 \pm 2 \cdot \frac{\hat{\sigma}_i^2}{\sqrt{T/2}}, \\
\hat{\sigma}_i \pm 2 \cdot \widehat{SE}(\hat{\sigma}_i) &= \hat{\sigma}_i \pm 2 \cdot \frac{\hat{\sigma}_i}{\sqrt{2T}} \\
\hat{\rho}_{ij} \pm 2 \cdot \widehat{SE}(\hat{\rho}_{ij}) &= \hat{\rho}_{ij} \pm 2 \cdot \frac{(1-\hat{\rho}_{ij}^2)}{\sqrt{T}}
\end{aligned}$$

Example 15 *To be completed*

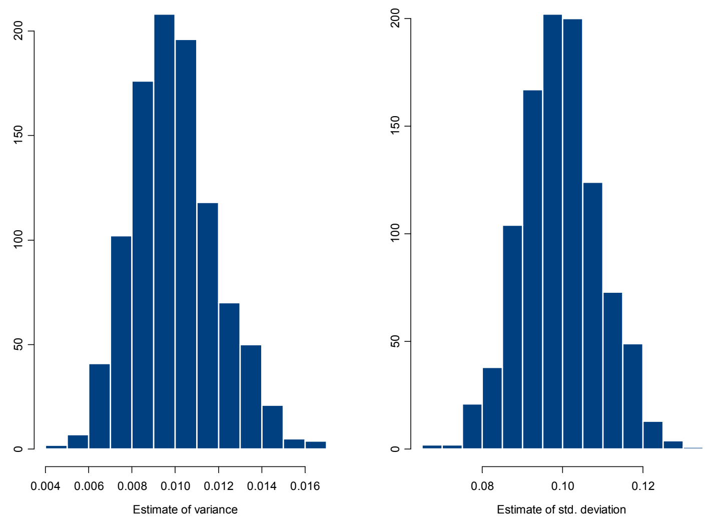


Figure 1.11: Histograms of $\hat{\sigma}^2$ and $\hat{\sigma}$ computed from $N = 1000$ Monte Carlo samples from CER model.

Evaluating the Statistical Properties of $\hat{\sigma}_i^2$, $\hat{\sigma}_i, \hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ by Monte Carlo simulation

We may evaluate the statistical properties of $\hat{\sigma}_i^2$, $\hat{\sigma}_i, \hat{\sigma}_{ij}$ and $\hat{\rho}_{ij}$ by Monte Carlo simulation in the same way that we evaluated the statistical properties of $\hat{\mu}_i$. Consider first the variability estimates $\hat{\sigma}_i^2$ and $\hat{\sigma}_i$. We use the simulation model (1.15) and $N = 1000$ simulated samples of size $T = 50$ to compute the estimates $\{\left(\hat{\sigma}^2\right)^1, \ldots, \left(\hat{\sigma}^2\right)^{1000}\}$ and $\{\hat{\sigma}^1, \ldots, \hat{\sigma}^{1000}\}$. The histograms of these values are displayed in figure 1.11.The histogram for the $\hat{\sigma}^2$ values is bell-shaped and slightly right skewed but is centered very close to $0.010 = \sigma^2$. The histogram for the $\hat{\sigma}$ values is more symmetric and is centered near $0.10 = \sigma$.

The average values of σ^2 and σ from the 1000 simulations are

$$
\begin{aligned}
\frac{1}{1000}\sum_{j=1}^{1000}\hat{\sigma}^2 &= 0.009952 \\
\frac{1}{1000}\sum_{j=1}^{1000}\hat{\sigma} &= 0.09928
\end{aligned}
$$

The sample standard deviation values of the Monte Carlo estimates of σ^2 and σ give approximations to $SE(\hat{\sigma}^2)$ and $SE(\hat{\sigma})$. Using the formulas (1.18) and (1.19) these values are

$$
\begin{aligned}
SE(\hat{\sigma}^2) &= \frac{(0.10)^2}{\sqrt{50/2}} = 0.002 \\
SE(\hat{\sigma}) &= \frac{0.10}{\sqrt{2 \cdot 50}} = 0.010
\end{aligned}
$$

1.4 Further Reading

To be completed

1.5 Appendix

1.5.1 Proofs of Some Technical Results

Result: $E[\hat{\mu}_i] = \mu_i$

Proof. Using the fact that $\hat{\mu}_i = T^{-1}\sum_{t=1}^{T} R_{it}$ and $R_{it} = \mu_i + \varepsilon_{it}$ we have

$$\begin{aligned}
E[\hat{\mu}_i] &= E\left[\frac{1}{T}\sum_{t=1}^{T} R_{it}\right] \\
&= E\left[\frac{1}{T}\sum_{t=1}^{T}(\mu_i + \varepsilon_{it})\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i + \frac{1}{T}\sum_{t=1}^{T}E[\varepsilon_{it}] \text{ (by the linearity of } E[\cdot]) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mu_i \text{ (since } E[\varepsilon_{it}] = 0,\ t = 1, \ldots, T) \\
&= \frac{1}{T}T \cdot \mu_i \\
&= \mu_i.
\end{aligned}$$

■

Result: $var(\mu_i) = \frac{\sigma_i^2}{T}$.

Proof. Using the fact that $\hat{\mu}_i = T^{-1}\sum_{t=1}^{T} R_{it}$ and $R_{it} = \mu_i + \varepsilon_{it}$ we have

$$\begin{aligned}
var(\hat{\mu}_i) &= var\left(\frac{1}{T}\sum_{t=1}^{T} R_{it}\right) \\
&= var\left(\frac{1}{T}\sum_{t=1}^{T}(\mu_i + \varepsilon_{it})\right) \text{ (in the CER model } R_{it} = \mu_i + \varepsilon_{it}) \\
&= var\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right) \text{ (since } \mu_i \text{ is a constant)} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}var(\varepsilon_{it}) \text{ (since } \varepsilon_{it} \text{ is independent over time)} \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sigma_i^2 \text{ (since } var(\varepsilon_{it}) = \sigma_i^2,\ t = 1, \ldots, T) \\
&= \frac{1}{T^2}T\sigma_i^2 \\
&= \frac{\sigma_i^2}{T}.
\end{aligned}$$

■

1.5.2 Some Special Probability Distributions Used in Statistical Inference

The Chi-Square distribution with T degrees of freedom

Let $Z_1, Z_2, \ldots, Z_T$ be independent standard normal random variables. That is,

$$Z_i \sim i.i.d.\ N(0,1),\ i = 1, \ldots, T.$$

Define a new random variable X such that

$$X = Z_1^2 + Z_2^2 + \cdots Z_T^2 = \sum_{i=1}^{T} Z_i^2.$$

Then X is a chi-square random variable with T degrees of freedom. Such a random variable is often denoted χ_T^2 and we use the notation $X \sim \chi_T^2$. The pdf of X is illustrated in Figure xxx for various values of T. Notice that X is only allowed to take non-negative values. The pdf is highly right skewed for small values of T and becomes symmetric as T gets large. Furthermore, it can be shown that

$$E[X] = T.$$

The chi-square distribution is used often in statistical inference and probabilities associated with chi-square random variables are needed. Critical values, which are just quantiles of the chi-square distribution, are used in typical calculations. To illustrate, suppose we wish to find the critical value of the chi-square distribution with T degrees of freedom such that the probability to the right of the critical value is α. Let $\chi_T^2(\alpha)$ denote this critical value[8]. Then

$$\Pr(X > \chi_T^2(\alpha)) = \alpha.$$

For example, if $T = 5$ and $\alpha = 0.05$ then $\chi_5^2(0.05) = 11.07$; if $T = 100$ then $\chi_{100}^2(0.05) = 124.34$.

1.5.3 Student's t distribution with T degrees of freedom

Let Z be a standard normal random variable, $Z \sim N(0,1)$, and let X be a chi-square random variable with T degrees of freedom, $X \sim \chi_T^2$. Assume

[8] Excel has functions for computing probabilities from the chi-square distribution.

that Z and X are independent. Define a new random variable t such that

$$t = \frac{Z}{\sqrt{X/T}}.$$

Then t is a Student's t random variable with T degrees of freedom and we use the notation $t \sim t_T$ to indicate that t is distributed Student-t. Figure xxx shows the pdf of t for various values of the degrees of freedom T. Notice that the pdf is symmetric about zero and has a bell shape like the normal. The tail thickness of the pdf is determined by the degrees of freedom. For small values of T, the tails are quite spread out and are thicker than the tails of the normal. As T gets large the tails shrink and become close to the normal. In fact, as $T \to \infty$ the pdf of the Student t converges to the pdf of the normal.

The Student-t distribution is used heavily in statistical inference and critical values from the distribution are often needed. Let $t_T(\alpha)$ denote the critical value such that

$$\Pr(t > t_T(\alpha)) = \alpha.$$

For example, if $T = 10$ and $\alpha = 0.025$ then $t_{10}(0.025) = 2.228$; if $T = 100$ then $t_{60}(0.025) = 2.00$. Since the Student-t distribution is symmetric about zero, we have that

$$\Pr(-t_T(\alpha) \leq t \leq t_T(\alpha)) = 1 - 2\alpha.$$

For example, if $T = 60$ and $\alpha = 2$ then $t_{60}(0.025) = 2$ and

$$\Pr(-t_{60}(0.025) \leq t \leq t_{60}(0.025)) = \Pr(-2 \leq t \leq 2) = 1 - 2(0.025) = 0.95.$$

1.6 Problems

To be completed

Bibliography

[1] Campbell, Lo and MacKinley (1998). *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.

Introduction to Financial Econometrics Appendix Matrix Algebra Review

Eric Zivot
Department of Economics
University of Washington

January 3, 2000
This version: February 6, 2001

1 Matrix Algebra Review

A *matrix* is just an array of numbers. The *dimension* of a matrix is determined by the number of its rows and columns. For example, a matrix $\mathbf{A}$ with n rows and m columns is illustrated below

$$\underset{(n\times m)}{\mathbf{A}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

where a_{ij} denotes the i^{th} row and j^{th} column element of $\mathbf{A}$.

A *vector* is simply a matrix with 1 column. For example,

$$\underset{(n\times 1)}{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is an $n \times 1$ vector with elements $x_1, x_2, \dots, x_n$. Vectors and matrices are often written in bold type (or underlined) to distinguish them from scalars (single elements of vectors or matrices).

The *transpose* of an $n \times m$ matrix $\mathbf{A}$ is a new matrix with the rows and columns of $\mathbf{A}$ interchanged and is denoted $\mathbf{A}'$ or $\mathbf{A}^{\intercal}$. For example,

$$\underset{(2\times 3)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \underset{(3\times 2)}{\mathbf{A}'} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

$$\underset{(3\times 1)}{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \underset{(1\times 3)}{\mathbf{x}'} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}.$$

A *symmetric* matrix $\mathbf{A}$ is such that $\mathbf{A} = \mathbf{A}'$. Obviously this can only occur if $\mathbf{A}$ is a *square* matrix; i.e., the number of rows of $\mathbf{A}$ is equal to the number of columns. For example, consider the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Clearly,

$$\mathbf{A}' = \mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

1.1 Basic Matrix Operations

1.1.1 Addition and subtraction

Matrix addition and subtraction are element by element operations and only apply to matrices of the same dimension. For example, let

$$\mathbf{A} = \begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}.$$

Then

$$\begin{aligned} \mathbf{A}+\mathbf{B} &= \begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 4+2 & 9+0 \\ 2+0 & 1+7 \end{bmatrix} = \begin{bmatrix} 6 & 9 \\ 2 & 8 \end{bmatrix}, \\ \mathbf{A}-\mathbf{B} &= \begin{bmatrix} 4 & 9 \\ 2 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} 4-2 & 9-0 \\ 2-0 & 1-7 \end{bmatrix} = \begin{bmatrix} 2 & 9 \\ 2 & -6 \end{bmatrix}. \end{aligned}$$

1.1.2 Scalar Multiplication

Here we refer to the multiplication of a matrix by a scalar number. This is also an element-by-element operation. For example, let $c = 2$ and

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ 0 & 5 \end{bmatrix}.$$

Then

$$c \cdot \mathbf{A} = \begin{bmatrix} 2\cdot 3 & 2\cdot(-1) \\ 2\cdot(0) & 2\cdot 5 \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ 0 & 10 \end{bmatrix}.$$

1.1.3 Matrix Multiplication

Matrix multiplication only applies to *conformable* matrices. $\mathbf{A}$ and $\mathbf{B}$ are conformable matrices of the number of columns in A is equal to the number of rows in $\mathbf{B}$. For example, if $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $m \times p$ then $\mathbf{A}$ and $\mathbf{B}$ are conformable. The mechanics of matrix multiplication is best explained by example. Let

$$\underset{(2\times 2)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } \underset{(2\times 3)}{\mathbf{B}} = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 2 \end{bmatrix}.$$

Then

$$\begin{aligned} \underset{(2\times 2)}{\mathbf{A}} \cdot \underset{(2\times 3)}{\mathbf{B}} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 3 & 4 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1\cdot 1 + 2\cdot 3 & 1\cdot 2 + 2\cdot 4 & 1\cdot 1 + 2\cdot 2 \\ 3\cdot 1 + 4\cdot 3 & 3\cdot 2 + 4\cdot 4 & 3\cdot 1 + 4\cdot 2 \end{bmatrix} \\ &= \begin{bmatrix} 7 & 10 & 5 \\ 15 & 22 & 11 \end{bmatrix} = \underset{(2\times 3)}{\mathbf{C}} \end{aligned}$$

The resulting matrix $\mathbf{C}$ has 2 rows and 3 columns. In general, if $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$ then $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$ is $n \times p$.

As another example, let

$$\underset{(2\times 2)}{\mathbf{A}} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ and } \underset{(2\times 1)}{\mathbf{B}} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}.$$

Then

$$\begin{aligned} \underset{(2\times 2)}{\mathbf{A}} \cdot \underset{(2\times 1)}{\mathbf{B}} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 6 \end{bmatrix} \\ &= \begin{bmatrix} 1\cdot 5 + 2\cdot 6 \\ 3\cdot 5 + 4\cdot 6 \end{bmatrix} \\ &= \begin{bmatrix} 17 \\ 39 \end{bmatrix}. \end{aligned}$$

As a &nal example, let

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}.$$

Then

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1\cdot 4 + 2\cdot 5 + 3\cdot 6 = 32$$

1.2 The Identity Matrix

The identity matrix plays a similar role as the number 1. Multiplying any number by 1 gives back that number. In matrix algebra, pre or post multiplying a matrix A by a conformable identity matrix gives back the matrix A. To illustrate, let

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

denote the 2 dimensional identity matrix and let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

denote an arbitrary 2×2 matrix. Then

$$\begin{aligned} \mathbf{I} \cdot \mathbf{A} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A} \end{aligned}$$

and

$$\begin{aligned} \mathbf{A} \cdot \mathbf{I} &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \mathbf{A}. \end{aligned}$$

1.3 Inverse Matrix

To be completed.

1.4 Representing Summation Using Vector Notation

Consider the sum

$$\sum_{k=1}^{n} x_k = x_1 + \cdots + x_k.$$

Let $\mathbf{x} = (x_1, \ldots, x_n)'$ be an $n \times 1$ vector and $\mathbf{1} = (1, \ldots, 1)'$ be an $n \times 1$ vector of ones. Then

$$\mathbf{x}'\mathbf{1} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = x_1 + \cdots + x_k = \sum_{k=1}^{n} x_k$$

and

$$\mathbf{1}'\mathbf{x} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 + \cdots + x_n = \sum_{k=1}^{n} x_k.$$

Next, consider the sum of squared x values

$$\sum_{k=1}^{n} x_k^2 = x_1^2 + \cdots + x_n^2.$$

This sum can be conveniently represented as

$$\mathbf{x}'\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1^2 + \cdots + x_n^2 = \sum_{k=1}^{n} x_k^2.$$

Last, consider the sum of cross products

$$\sum_{k=1}^{n} x_k y_k = x_1 y_1 + \cdots x_n y_n.$$

This sum can be compactly represented by

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + \cdots x_n y_n = \sum_{k=1}^{n} x_k y_k.$$

Note that $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$.

1.5 Representing Systems of Linear Equations Using Matrix Algebra

Consider the system of two linear equations

$$\begin{aligned} x + y &= 1 \qquad (1) \\ 2x - y &= 1 \qquad (2) \end{aligned}$$

which is illustrated in Figure xxx. Equations (1) and (2) represent two straight lines which intersect at the point $x = \frac{2}{3}$ and $y = \frac{1}{3}$. This point of intersection is determined by solving for the values of x and y such that $x + y = 2x - y$[1].

[1]Soving for x gives $x = 2y$. Substituting this value into the equation $x + y = 1$ gives $2y + y = 1$ and solving for y gives $y = 1/3$. Solving for x then gives $x = 2/3$.

The two linear equations can be written in matrix form as

$$\begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

or

$$\mathbf{A} \cdot \mathbf{z} = \mathbf{b}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}, \ \mathbf{z} = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

If there was a (2×2) matrix $\mathbf{B}$, with elements b_{ij}, such that $\mathbf{B} \cdot \mathbf{A} = \mathbf{I}$, where $\mathbf{I}$ is the (2×2) identity matrix, then we could solve for the elements in $\mathbf{z}$ as follows. In the equation $\mathbf{A} \cdot \mathbf{z} = \mathbf{b}$, pre-multiply both sides by $\mathbf{B}$ to give

$$\begin{aligned} \mathbf{B} \cdot \mathbf{A} \cdot \mathbf{z} &= \mathbf{B} \cdot \mathbf{b} \\ &\Longrightarrow \mathbf{I} \cdot \mathbf{z} = \mathbf{B} \cdot \mathbf{b} \\ &\Longrightarrow \mathbf{z} = \mathbf{B} \cdot \mathbf{b} \end{aligned}$$

or

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} b_{11} \cdot 1 + b_{12} \cdot 1 \\ b_{21} \cdot 1 + b_{22} \cdot 1 \end{bmatrix}$$

If such a matrix $\mathbf{B}$ exists it is called the inverse of $\mathbf{A}$ and is denoted $\mathbf{A}^{-1}$. Intuitively, the inverse matrix $\mathbf{A}^{-1}$ plays a similar role as the inverse of a number. Suppose a is a number; e.g., $a = 2$. Then we know that $\frac{1}{a} \cdot a = a^{-1}a = 1$. Similarly, in matrix algebra $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. Next, consider solving the equation $ax = 1$. By simple division we have that $x = \frac{1}{a}x = a^{-1}x$. Similarly, in matrix algebra if we want to solve the system of equation $\mathbf{Ax} = \mathbf{b}$ we multiply by $\mathbf{A}^{-1}$ and get $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Using $\mathbf{B} = \mathbf{A}^{-1}$, we may express the solution for $\mathbf{z}$ as

$$\mathbf{z} = \mathbf{A}^{-1}\mathbf{b}.$$

As long as we can determine the elements in $\mathbf{A}^{-1}$ then we can solve for the values of x and y in the vector $\mathbf{z}$. Since the system of linear equations has a solution as long as the two lines intersect, we can determine the elements in $\mathbf{A}^{-1}$ provided the two lines are not parallel.

There are general numerical algorithms for &nding the elements of $\mathbf{A}^{-1}$ and typical spreadsheet programs like Excel have these algorithms available. However, if $\mathbf{A}$ is a (2×2) matrix then there is a simple formula for $\mathbf{A}^{-1}$. Let $\mathbf{A}$ be a (2×2) matrix such that

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Then

$$\mathbf{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

By brute force matrix multiplication we can verify this formula

$$\begin{aligned}
\mathbf{A}^{-1}\mathbf{A} &= \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \\
&= \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22}a_{11} - a_{12}a_{21} & a_{22}a_{12} - a_{12}a_{22} \\ -a_{21}a_{11} + a_{11}a_{21} & -a_{21}a_{12} + a_{11}a_{22} \end{bmatrix} \\
&= \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22}a_{11} - a_{12}a_{21} & 0 \\ 0 & -a_{21}a_{12} + a_{11}a_{22} \end{bmatrix} \\
&= \begin{bmatrix} \frac{a_{22}a_{11} - a_{12}a_{21}}{a_{11}a_{22} - a_{21}a_{12}} & 0 \\ 0 & \frac{-a_{21}a_{12} + a_{11}a_{22}}{a_{11}a_{22} - a_{21}a_{12}} \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.
\end{aligned}$$

Let's apply the above rule to find the inverse of $\mathbf{A}$ in our example:

$$\mathbf{A}^{-1} = \frac{1}{-1-2} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{-1}{3} \end{bmatrix}.$$

Notice that

$$\mathbf{A}^{-1}\mathbf{A} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{-1}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Our solution for z is then

$$\begin{aligned}
\mathbf{z} &= \mathbf{A}^{-1}\mathbf{b} \\
&= \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{-1}{3} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}
\end{aligned}$$

so that $x = \frac{2}{3}$ and $y = \frac{1}{3}$.

In general, if we have n linear equations in n unknown variables we may write the system of equations as

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\vdots &= \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n
\end{aligned}$$

which we may then express in matrix form as

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

or

$$\underset{(n\times n)}{\mathbf{A}} \cdot \underset{(n\times 1)}{\mathbf{x}} = \underset{(n\times 1)}{\mathbf{b}}.$$

The solution to the system of equations is given by

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

where $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{I}$ is the $(n \times n)$ identity matrix. If the number of equations is greater than two, then we generally use numerical algorithms to &nd the elements in $\mathbf{A}^{-1}$.

2 Further Reading

Excellent treatments of portfolio theory using matrix algebra are given in Ingersol (1987), Huang and Litzenberger (1988) and Campbell, Lo and MacKinlay (1996).

3 Problems

To be completed

References

[1] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Priceton, New Jersey: Princeton University Press.

[2] Huang, C.-F., and Litzenbeger, R.H. (1988). *Foundations for Financial Economics*. New York: North-Holland.

[3] Ingersoll, J.E. (1987). *Theory of Financial Decision Making*. Totowa, New Jersey: Rowman & Little&eld.

University of Washington
Department of Economics

Winter 1999
Eric Zivot

Economics 483

Final Exam

This is a closed book and closed note exam. However, you are allowed one page of handwritten notes. Answer all questions and write all answers in a blue book. Total points = 100.

I. Portfolio Theory (20 points)

1. Consider the problem of allocating wealth between a collection of N risky assets and a risk-free asset (T-bill) under the assumption that investors only care about maximizing portfolio expected return and minimizing portfolio variance. Use the graph below to answer the following questions.

a. Mark on the graph the set of efficient portfolios for the risky assets only (transfer the graph to your blue book). Briefly describe how you would compute this set using Excel. (10 pts)

b. Mark on the graph the set of efficient portfolios that include risky assets and a single risk-free asset (transfer the graph to your blue book). Briefly describe how you would compute this set using Excel. (10 pts)

II. CAPM (20 points)

1. Consider the CAPM regression

$$R_t - r_f = \alpha + \beta(R_{Mt} - r_f) + \varepsilon_t,\ t = 1,\ldots,T$$
$$\varepsilon_t \sim iid\,\mathrm{N}(0, \sigma_\varepsilon^2) \text{ and } R_{Mt} \text{ is independent of } \varepsilon_t \text{ for all } t$$

where R_t denotes the return on an asset or portfolio, R_{Mt} denotes the return on the market portfolio proxy and r_f denotes the risk-free T-bill rate. Let μ and μ_M denote the expected returns on the asset and the market, respectively, and let σ^2 and σ_M^2 denote the variances of the asset and the market, respectively. Finally, let σ_{RM} denote the covariance between the asset and the market.

a. What is the interpretation of α and β in the CAPM regression? What restriction does the CAPM place on the value of α? (4 pts)

b. What is the interpretation of ε_t in the CAPM regression? (2 pts)

c. Using the CAPM regression compute $E[R_t]$ and $\text{var}(R_t)$. (2 pts)

d. Using the expression for $\text{var}(R_t)$, what is the proportion of the variance of the asset due to the variability in the market return and what is the proportion unexplained by variability in the market? (2 pts)

2. The following output is based on estimating the CAPM regression for IBM and an equally weighted portfolio of 15 stocks using monthly return data over the period January 1978 to December 1982:

$$R_{IBM} - r_f = \underset{(0.0068)}{-0.0002} + \underset{(0.0888)}{0.3390}*(R_M - r_f),\ R^2 = 0.2008,\ \text{var}(\varepsilon_{IBM}) = (0.0524)^2$$

$$R_{port} - r_f = \underset{(0.0030)}{0.0006} + \underset{(0.0447)}{0.6316}*(R_M - r_f),\ R^2 = 0.6280,\ \text{var}(\varepsilon_{port}) = (0.0335)^2$$

a. For IBM and the portfolio of 15 stocks, what are the estimated values of α and β and what are the estimated standard errors for these estimates? (2 pts)

b. Is the beta for the portfolio estimated more precisely than the beta for IBM? Why or why not? (2 pts)

c. For each regression, what is the proportion of market or systematic risk and what is the proportion of firm specific or unsystematic risk? Why should the portfolio have a greater proportion of systematic risk and smaller value of $SD(\varepsilon)$ than IBM? (2 pts)

d. Based on the regression estimates, does the CAPM appear to hold for IBM and the portfolio? Justify your answer. (4 pts)

III. Return Calculations (20 points)

1. Consider a portfolio of 3 risky stocks denoted by A, B and C (say Apple, Boeing and Coca Cola). Let R_A, R_B and R_C denote the monthly returns on these stock and it is assumed that these returns are jointly normally distributed with means μ_i (i = A,B,C), variances σ_i^2 (i = A,B,C) and covariances σ_{ij} (i = A,B,C and $i \neq j$). Consider forming a portfolio of these stocks where x_i = share of wealth invested in asset i such that $x_A + x_B + x_C = 1$.

a. What is the expected return on the portfolio? (2 pts)
b. What is the variance of the portfolio return? (2 pts)
c. What is the probability distribution for the portfolio return? (2 pts)

2. Throughout the course we have made the assumption that the continuously compounded returns on risky assets (e.g. stocks) are normally distributed. Based on the data analysis we have done in the labs and in class, is this a believable assumption? Briefly justify your answer. (8 pts)

3. Consider the following monthly data for Microsoft stock over the period December 1995 through December 1996:

End of Month Price Data for Microsoft Stock	
December, 1995	43.12
January, 1996	43.87
February, 1996	47.06
March, 1996	47.75
April, 1996	51.37
May, 1996	57.56
June, 1996	59.19
July, 1996	61.16
August, 1996	60.31
September, 1996	61.25
October, 1996	66.06
November, 1996	68.69
December, 1996	78.87

a. Using the data in the table, what is the continuously compounded monthly return between December, 1995 and January 1996? (2 pts)

b. Assuming that the continuously compounded monthly return you computed in part (a) is the same for 12 months, what is the continuously compounded annual return? (2 pts)

c. Using the data in the table, compute the actual (annual) continuously compounded return between December 1995 and December 1996. Compare with your result in part (b). (2 pts)

IV. Arbitrage (15 points)

a. What is an *arbitrage opportunity*? (5 pts)

b. Give a simple example of an arbitrage opportunity. (5 pts)