CHAPTER 12

STATISTICS FOR PORTFOLIOS

OVERVIEW

In order to understand and work through Chapters 13–16, you will need to know some statistics. If you're like a lot of finance students, you've had a statistics course and forgotten much of what you learned there. This chapter is a refresher—it show you exactly what you need in order to proceed with the succeeding chapters, using Excel to do all the calculations. (Excel is a great statistical toolbox—someday all business-school statistics courses will use it. In the meantime you're stuck with this chapter.)

Finance Concepts Discussed

- How to calculate stock returns and adjust them for dividends and stock splits
- Return mean, variance, and standard deviation for an asset
- Return mean and variance for a portfolio of two assets
- Regressions

Excel Functions Used

- **Average**
- **Var** and **Varp**
- **Stdev** and **Stdevp**
- **Covar** and **Correl**
- **Trendlines** (Excel's term for regressions)
- **Slope, Intercept, Rsq**

12.1 Basic Statistics for Asset Returns: Mean, Standard Deviation, Covariance, and Correlation

In this section you learn to calculate the return on a stock and its statistics: the *mean* (interchangeably referred to as the *average* or *expected* return), the *variance,* and the *standard deviation*.

General Motors Stock and Its Returns

The following spreadsheet shows data for General Motors (GM) stock during the decade of the 1990s. For each year, we've given the closing price of GM stock and the dividend the company paid during the year.[1] We've also calculated the annual returns and their statistics; these calculations are explained after the table:

[1]The "closing price" is the price of the stock at the end of the day.

	A	B	C	D	E
1	**PRICE AND DIVIDEND DATA FOR GENERAL MOTORS (GM)**				
2	**Date**	**Closing Price**	**Dividend**	**Annual return**	
3	29-Dec-89	42.2500	-		
4	31-Dec-90	34.3750	3.00	-11.54%	<-- =(C4+B4)/B3-1
5	31-Dec-91	28.8750	1.60	-11.35%	<-- =(C5+B5)/B4-1
6	31-Dec-92	32.2500	1.40	16.54%	
7	31-Dec-93	54.8750	0.80	72.64%	
8	30-Dec-94	42.1250	0.80	-21.78%	
9	29-Dec-95	52.8750	1.10	28.13%	
10	31-Dec-96	55.7500	1.60	8.46%	
11	31-Dec-97	60.7500	5.59	19.00%	
12	31-Dec-98	71.5625	2.00	21.09%	
13	31-Dec-99	72.6875	14.15	21.34%	
14					
15	Average return, $E(r_{GM})$			14.25%	<-- =AVERAGE(D4:D13)
16	Variance of return, σ^2_{GM}			0.0638	<-- =VARP(D4:D13)
17	Standard deviation of return, σ_{GM}			25.25%	<-- =STDEVP(D4:D13)

Suppose you had bought a share of GM at the end of December 1989 for $42.25 and sold it a year later, at the end of December 1990, for $34.375. During this year, GM paid a per-share dividend of $3.[2] Your return from holding GM throughout 1990 would have been

$$r_{GM,1990} = \frac{P_{GM,1990} + Div_{GM,1990} - P_{GM,1989}}{P_{GM,1989}} = \frac{34.375 + 3.00 - 42.25}{42.25} = -11.54\%$$

Several notes:

- We use $r_{GM,1990}$ to denote the return on GM stock in 1990 and we use $Div_{GM,1990}$ to denote GM's dividend in 1990.
- The numerator of $r_{GM,1990}$ is

  $$P_{GM,1990} + Div_{GM,1990} - P_{GM,1989} = 34.375 + 3.00 - 42.25 = -4.875$$

  This is the gain on holding GM during the year (in this case it's a negative "gain": a loss of $4.875). The denominator of $r_{GM,1990}$ is the initial investment from buying GM stock at the beginning of the year.
- In cell D4 of the spreadsheet we've written $r_{GM,1990}$, the return for 1990, in a slightly different form as **(C4+B4)/B3−1:**

  $$r_{GM,1990} = \frac{P_{GM,1990} + Div_{GM,1990} - P_{GM,1989}}{P_{GM,1989}} = \frac{P_{GM,1990} + Div_{GM,1990}}{P_{GM,1989}} - 1$$

  Cells D15, D16, and D17 give the return statistics for GM.
- Cell D15: The average return over the decade is 14.25% per year. This number is also called the *mean return* and it's calculated with the Excel function =**Average(D4:D13)**. We often use the past returns to predict future returns. When we make this use of the data, we also call the mean the *expected return*, meaning that we use the historic average of GM's stock returns as a prediction of what the stock will return in the future. We sometimes use the notations.

---

[2]Actually, the company paid four quarterly dividends of $0.75, but we've added these together to get the annual dividend.

$E(r_{GM})$ or $\bar{r}_{GM}$. In this book, the terms mean, average, and expected return are used almost interchangeably. The formal definition is

$$mean\ GM\ return = E\left(r_{GM}\right) = \bar{r}_{GM} = \frac{r_{GM,1990} + r_{GM,1991} + \cdots + r_{GM,1999}}{10}$$

You might wonder at the number of expressions (mean, average, expected return) and the number of symbols ($E(r_{GM}), \bar{r}_{GM}$) for the same idea. We introduce them all for convenience and because, in your further finance studies, you're likely to see them used synonymously.

- Cell D16: The variance of the annual returns is 6.38%. Variance and standard deviation are statistical measures of the variability of the returns. The variance is calculated with the Excel function **=Varp(D4:D13)**. (See the Excel Note box further on for more information about this function and its cousin, **=Var(D4:D13)**.) The variance is often denoted by the Greek symbol σ^2_{GM} (pronounced "sigma squared of GM"); sometimes it's written as $Var(r_{GM})$. The formal definition of the variance is

$$Var\left(r_{GM}\right) = \sigma^2_{GM} = \frac{\left(r_{GM,1990} - \bar{r}_{GM}\right)^2 + \left(r_{GM,1991} - \bar{r}_{GM}\right)^2 + \cdots + \left(r_{GM,1999} - \bar{r}_{GM}\right)^2}{10}$$

- Cell D17: The standard deviation of the annual returns is the square root of the variance: $\sqrt{0.0638} = 25.25\%$. Excel has two functions, **Stdevp** and **Stdev**, to do this calculation directly. Since we usually use **Varp** for the variance, we will use **Stdevp** for the standard deviation. It is common to use the Greek letter sigma for the standard deviation, writing σ_{GM} (pronounced "sigma of GM").

EXCEL NOTE

EXCEL AND STATISTICS (SKIP UNTIL LATER, OR PERHAPS FOREVER, IF YOU LIKE)

Excel has two variance functions, **Varp** and **Var**. The former measures the "population variance," and the latter measures the "sample variance." Similarly, Excel has two functions for the standard deviation, **Stdevp** and **Stdev**. In this book we use only the functions **Varp** and **Stdevp**. This box is a reminder but not an explanation of the difference between the two concepts.

If you have return data $\{r_{stock,1}, r_{stock,2}, \ldots, r_{stock,N}\}$ for a *stock,* then the mean return is

$$\bar{r}_{stock} = \frac{1}{N}\sum_{t=1}^{N} r_{stock,t}$$

The definitions of the two variance functions are

$$Varp(\{r_{stock,1}, r_{stock,2}, \ldots, r_{stock,N}\}) = \frac{1}{N}\sum_{j=1}^{N}(r_{stock,j} - \bar{r}_i)^2$$

$$Var(\{r_{stock,1}, r_{stock,2}, \ldots, r_{stock,N}\}) = \frac{1}{N-1}\sum_{j=1}^{N}(r_{stock,j} - \bar{r}_i)^2$$

There's a long story about the difference between these two concepts, which we leave for someone else (like your statistics instructor) to explain. Suffice it to say that in the examples covered in this book we use **Varp** and its standard deviation equivalent **Stdevp.**

Finally, you might wonder why there are two expressions—the variance and the standard deviation—that measure the variability. The answer has to do with the units of these expressions. Each term in the variance is squared in order to make everything positive. But this means that the units of the variance are "percent squared," which is a bit difficult to understand. The standard deviation, the square root of the variance, reduces the squared percentages of the variance back to "percent." This way the mean and the standard deviation have the same units.

Microsoft Stock and Its Returns

The GM example earlier illustrated the adjustment of the stock return data to include dividends. We now use Microsoft stock to show you how stock returns are affected by stock splits. A *stock split* occurs when stockholders get multiple shares of stock for each share they own. The most typical stock split is a "2-for-1" split, in which shareholders get one additional share for each share they already own (see Figure 12.1 for a Microsoft stock split announcement in 1996).

Figure 12.1 On 12 November 1996, Microsoft announced a 2-for-1 stock split. Shareholders owning one share on 22 November 1996 would be mailed an additional share of stock. This increased the number of shares of the company from 600 million to 1.2 billion. The Microsoft statement hints at the company's reasoning for the split: With its stock trading at almost \$150 per share before the split, Microsoft used the split to reduce the price of the share in order to put it into a range that would make it "more accessible to a broader base of investors."

Microsoft (MSFT) paid no dividends in the 1990–1999 decade, but the stock split several times. Here are some data:

	A	B	C
1	**PRICE AND STOCK SPLIT DATA FOR MICROSOFT (MSFT)**		
2	**Date**	**Closing Price**	**Stock split during year?**
3	29-Dec-89	87.0000	
4	31-Dec-90	75.2500	2.0 for 1
5	31-Dec-91	111.2500	1.5 for 1
6	31-Dec-92	85.3750	1.5 for 1
7	31-Dec-93	80.6250	no
8	30-Dec-94	61.1250	2.0 for 1
9	29-Dec-95	87.7500	no
10	31-Dec-96	82.6250	2.0 for 1
11	31-Dec-97	129.2500	no
12	31-Dec-98	138.6875	2.0 for 1
13	31-Dec-99	116.7500	2.0 for 1

Here's what these stock splits mean for Microsoft shareholders: Suppose you had bought one share of MSFT on 29 December 1989 for $87.00 and held it throughout 1990. During 1990, Microsoft *split* its stock, giving shareholders two shares for every one share they owned. At the end of 1990, each of these (split) shares was worth $75.25, so that your $87 investment had grown to $150.25! The return for the year is therefore

$$\begin{aligned} r_{MSFT,1990} &= \frac{(P_{MSFT,31Dec90}) * 2}{P_{MSFT,29Dec89}} - 1 \\ &= \frac{150.50}{87} - 1 = 72.99\% \end{aligned}$$

The "2" in the formula above is the stock split *adjustment factor,* which shows that one share of Microsoft owned at the beginning of 1990 became two shares by the end of the year. In the spreadsheet below we calculate the *cumulative adjustment factor*. This shows you how your end-1989 $87.00 investment in MSFT would have grown throughout the decade if you correctly account for the stock splits.

	A	B	C	D	E	F	G
16	Date	Closing Price	Stock split during year?	Cumulative adjustment factor	Adjusted price	Annual return	
17	29-Dec-89	87.0000		1	87.00		
18	31-Dec-90	75.2500	2.0 for 1	2	150.50	72.99%	<-- =E18/E17-1
19	31-Dec-91	111.2500	1.5 for 1	3	333.75	121.76%	<-- =E19/E18-1
20	31-Dec-92	85.3750	1.5 for 1	4.5	384.19	15.11%	
21	31-Dec-93	80.6250	no	4.5	362.81	-5.56%	
22	30-Dec-94	61.1250	2.0 for 1	9	550.13	51.63%	
23	29-Dec-95	87.7500	no	9	789.75	43.56%	
24	31-Dec-96	82.6250	2.0 for 1	18	1,487.25	88.32%	
25	31-Dec-97	129.2500	no	18	2,326.50	56.43%	
26	31-Dec-98	138.6875	2.0 for 1	36	4,992.75	114.60%	
27	31-Dec-99	116.7500	2.0 for 1	72	8,406.00	68.36%	
28							
29	Average return, $E(r_{MSFT})$					62.72%	<-- =AVERAGE(F18:F27)
30	Variance of return, σ^2_{MSFT}					14.43%	<-- =VARP(F18:F27)
31	Standard deviation of return, σ_{MSFT}					37.99%	<-- =SQRT(F30)
32							
33							

The cumulative adjustment factor is the product of all the splits: 72 = 2*1.5*1.5*2*2*2

Taking into account the stock splits, your $87.00 investment in MSFT would have grown to $8,406 by the end of the decade! During the 1990s, MSFT gave an average annual return of 62.72%; this return had a standard deviation of 37.99%.[3]

STOCK SPLITS AND THE CUMULATIVE ADJUSTMENT FACTOR

On 31 January 2002, you bought one share of XYZ stock for $50. One year minus one day later, on 30 January 2003, your share of XYZ stock is trading at $80. At the end of the day the stock *splits:* For every share you own, you now have *two* shares. In a logical world, this would mean that the price of the share should fall by 50%, so that on 31 January 2003, XYZ trades at $40 per share.[4]

Now suppose you're trying to calculate your return on the stock. Is it $\$40/\$50 - 1 = -20\%$ or is the return $(2 * \$40)/\$50 - 1 = 60\%$? The latter, of course! *You adjusted the stock price by the adjustment factor*.

Suppose that in July 2003 XYZ splits 1.5 for 1 and that on 31 January 2004 the price per share is $25. Then your return since you bought the stock is

$$\frac{2 * 1.5 * \$25}{\$50} - 1 = \frac{3 * \$25}{\$50} - 1 = 50\%$$

The cumulative adjustment factor is the product of all the splits since you bought the stock.

---

[3]Adding or subtracting the standard deviation from the average gives a plausible range for Microsoft stock returns. Roughly speaking, the 37.99% standard deviation indicates that with a 68% probability, the Microsoft stock returns are in the range between 24.73% and 100.71%. These two numbers are computed by $24.73\% = 62.72\% - 37.99\%$ and $100.71\% = 62.72\% + 37.99\%$.

[4]The world is not all that logical, but this in fact usually happens—when a stock splits 2 for 1, its post-split price is usually very close to half its pre-split price. If the stock splits on a 1.5 for 1 basis, the post-split price is close to two-thirds $(2/3 = 1/1.5)$ its pre-split price.

12.2 Downloaded Data From Commercial Sources Is Adjusted for Dividends and Splits

The author's two favorite data sources for information about stock prices, dividends, and stock splits are Yahoo, which is free, and the Center for Research in Security Prices (CRSP), a database that originates from the University of Chicago (many universities subscribe to CRSP—ask your data manager).[5] When you download data from these sources, they automatically adjust the price data to account for dividends and splits. So you don't have to do all the adjustment calculations illustrated in the previous section.[6]

It is important to note, however, that the adjustments made by Yahoo and CRSP may look different from the ones we made above. For example, here's the adjusted Microsoft data from Yahoo:

	A	B	C	D
1	**DOWNLOADED ADJUSTED DATA FROM YAHOO FOR MICROSOFT**			
2	**Date**	**MSFT adjusted price**		
3	29-Dec-89	1.2083		
4	31-Dec-90	2.0903	73.00%	<-- =B4/B3-1
5	31-Dec-91	4.6354	121.76%	<-- =B5/B4-1
6	31-Dec-92	5.3359	15.11%	<-- =B6/B5-1
7	31-Dec-93	5.0391	-5.56%	<-- =B7/B6-1
8	30-Dec-94	7.6406	51.63%	<-- =B8/B7-1
9	29-Dec-95	10.9688	43.56%	<-- =B9/B8-1
10	31-Dec-96	20.6562	88.32%	<-- =B10/B9-1
11	31-Dec-97	32.3125	56.43%	<-- =B11/B10-1
12	31-Dec-98	69.3438	114.60%	<-- =B12/B11-1
13	31-Dec-99	116.7500	68.36%	<-- =B13/B12-1
14				
15	Average return, $E(r_{MSFT})$		62.72%	<-- =AVERAGE(C4:C13)
16	Variance of return, σ^2_{MSFT}		14.43%	<-- =VARP(C4:C13)
17	Standard deviation of return, σ_{MSFT}		37.99%	<-- =STDEVP(C4:C13)

The annual return statistics are the same, but the method of price adjustment is different: Yahoo has adjusted the stock prices so that the stock price on the last date ($116.75) is the same as the market price on that date. All previous prices have been adjusted accordingly. For example, the 29 December 1989 Yahoo price for MSFT of $1.2083 is 1/72 times the actual market price on this date—this adjustment is made since the stock split 72 times in the period 1990–1999.

---

[5]For penniless students, Yahoo is especially useful. Appendix 12.1 shows you how to download financial data from Yahoo.

[6]If it's all in the downloaded data, why the heck did we do all the work in this section? The answer, of course, is that it helps to understand what the numbers are telling you.

THE BOTTOM LINE ON DOWNLOADED DATA

Don't worry too much about how the adjustment is done. Calculate your returns from the adjusted stock price data given by your data provider. They usually do the corrections right.

12.3 Covariance and Correlation: Two Additional Statistics

So far we've looked at statistics—mean, variance, standard deviation—that relate to the returns of an individual stock. In this section we examine two statistics—*covariance* and *correlation*—that relate the returns of two stocks to each other. We continue to use our data for GM and MSFT. In the following spreadsheet, we've put the returns for both stocks on one spreadsheet and calculated the covariance and correlation (cells B17:B19):

	A	B	C	D
1	**GM AND MSFT, ANNUAL RETURN DATA**			
2	**Date**	**GM return**	**MSFT return**	
3	31-Dec-90	-11.54%	72.99%	
4	31-Dec-91	-11.35%	121.76%	
5	31-Dec-92	16.54%	15.11%	
6	31-Dec-93	72.64%	-5.56%	
7	30-Dec-94	-21.78%	51.63%	
8	29-Dec-95	28.13%	43.56%	
9	31-Dec-96	8.46%	88.32%	
10	31-Dec-97	19.00%	56.43%	
11	31-Dec-98	21.09%	114.60%	
12	31-Dec-99	21.34%	68.36%	
13				
14	Average return, $E(r_{GM})$ and $E(r_{MSFT})$	14.25%	62.72%	
15	Variance of return, σ^2_{GM} and σ^2_{MSFT}	6.38%	14.43%	
16	Standard deviation of return, σ_{GM} and σ_{MSFT}	25.25%	37.99%	
17	Covariance of returns, $Cov(r_{GM},r_{MSFT})$	-0.0552		<-- =COVAR(B3:B12,C3:C12)
18	Correlation of returns, $\rho_{GM,MSFT}$	-0.5755		<-- =CORREL(B3:B12,C3:C12)
19		-0.5755		<-- =B17/(B16*C16)

The *covariance* between two series is a measure of how much the series (in our case, the returns on GM and MSFT) move up or down together. The formal definition is

$$\begin{aligned} Cov(r_{GM}, r_{MSFT}) &= \sigma_{GM,MSFT} \\ &= \frac{1}{10}\left\{ \begin{array}{l} (r_{GM,1} - \bar{r}_{GM})(r_{MSFT,1} - \bar{r}_{MSFT}) + (r_{GM,2} - \bar{r}_{GM})(r_{MSFT,2} - \bar{r}_{MSFT}) \\ + \cdots + (r_{GM,10} - \bar{r}_{GM})(r_{MSFT,10} - \bar{r}_{MSFT}) \end{array} \right\} \end{aligned}$$

The idea behind the formula is to measure the deviations of each data point from its average and to multiply these deviations. As you can see from cell B17, Excel has a function **Covar**, which, when applied directly to the returns in columns B and C, calculates the covariance. Notice that the covariance is sometimes written as $\sigma_{GM,MSFT}$.

Calculating the covariance the long way using the formal definition may give you some more insight into what the covariance measures and what Excel's **Covar** function does.

	A	B	C	D	E	F	G	H
1	**CALCULATING THE COVARIANCE THE LONG TEDIOUS WAY**							
2	**Date**	**GM return**	**MSFT return**		**GM return minus average**	**MSFT return minus average**	=C3-$C$14 **Product**	
3	31-Dec-90	-11.54%	72.99%	=B3-$B$14-->	-25.79%	10.27%	-0.0265	<-- =E3*F3
4	31-Dec-91	-11.35%	121.76%		-25.60%	59.04%	-0.1511	
5	31-Dec-92	16.54%	15.11%		2.28%	-47.61%	-0.0109	
6	31-Dec-93	72.64%	-5.56%		58.38%	-68.28%	-0.3987	
7	30-Dec-94	-21.78%	51.63%		-36.03%	-11.09%	0.0400	
8	29-Dec-95	28.13%	43.56%		13.88%	-19.16%	-0.0266	
9	31-Dec-96	8.46%	88.32%		-5.79%	25.60%	-0.0148	
10	31-Dec-97	19.00%	56.43%		4.74%	-6.29%	-0.0030	
11	31-Dec-98	21.09%	114.60%		6.84%	51.88%	0.0355	
12	31-Dec-99	21.34%	68.36%		7.09%	5.64%	0.0040	
13								
14	Average return	14.25%	62.72%	<-- =AVERAGE(C3:C12)		Covariance	-0.0552	<-- =AVERAGE(G3:G12)
15						Covariance	-0.0552	<-- =COVAR(B3:B12,C3:C12)
16						Correlation	-0.5755	<-- =CORREL(B3:B12,C3:C12)
17						Correlation	-0.5755	<-- =G14/(STDEVP(B3:B12)*STDEVP(C3:C12))

In cell E3, we've subtracted GM's 1990 return of −11.54% from its decade average return of 14.25% (cell B14); the resulting number indicates that in 1990 GM stock underperformed its average by −25.79%. During the same year, MSFT overperformed its average by 10.27%. The covariance takes the product of these two numbers ($-25.79\% * 10.27\% = -0.0265$) and similar numbers for each of the other years and averages them (cell G14). As you can see, Excel's **Covar** function gives the same result (cell G15) and saves a lot of work. The covariance of −0.0552 for GM and MSFT tells us that, on average, when GM exceeded its mean, MSFT was below its mean, and vice versa.

Another common measure of how much two data series move up or down together is the *correlation coefficient*. The correlation coefficient is always between −1 and +1, which—as you'll see in the next subsection—makes it possible for us to be more precise about how the two sets of returns move together. Roughly speaking, two sets of returns that have a correlation coefficient of −1 vary *perfectly inversely,* by which we mean that when one return goes up (or down), we can perfectly predict how the other return goes down (or up). A correlation coefficient of +1 means that the returns vary in *perfect tandem,* by which we mean that when one return goes up (or down), we can perfectly predict how the other return goes up (or down). A correlation coefficient between −1 and +1 means that the two sets of returns vary together less than perfectly.

The correlation coefficient is defined as

$$Correlation(r_{GM}, r_{MSFT}) = \rho_{GM,MSFT} = \frac{Cov(r_{GM}, r_{MSFT})}{\sigma_{GM}\sigma_{MSFT}}$$

The Greek letter ρ (pronounced "rho") is often used as a symbol for the correlation coefficient. In the spreadsheet above, we calculate the correlation coefficient in two ways: In cell G16 of the previous spreadsheet, we use the Excel function **Correl** to compute the correlation. Cell G17 applies the formula $Cov(r_{GM}, r_{MSFT})/\sigma_{GM}\sigma_{MSFT}$ (and, of course, gets the same result).

Some Facts About Covariance and Correlation

Here are some facts about covariance and correlation. We state them without much attempt at elaborate explanation or proof.

Fact 1: Covariance is affected by units; correlation isn't. Here's an example. In the spreadsheet below, we've presented the annual returns as whole numbers instead of percentages (writing GM's 1990 return as −11.54 instead of −11.54%). The covariance (cell B18) is now −552.10, which is 10,000 times our previous calculation. But the correlation coefficient (B19) remains the same as before, −0.5755.

	A	B	C	D
1	**GM AND MSFT, ANNUAL RETURN DATA Percentages presented as whole numbers**			
2	**Date**	**GM annual return**	**MSFT annual return**	
3	29-Dec-89			
4	31-Dec-90	-11.54	72.99	
5	31-Dec-91	-11.35	121.76	
6	31-Dec-92	16.54	15.11	
7	31-Dec-93	72.64	-5.56	
8	30-Dec-94	-21.78	51.63	
9	29-Dec-95	28.13	43.56	
10	31-Dec-96	8.46	88.32	
11	31-Dec-97	19.00	56.43	
12	31-Dec-98	21.09	114.60	
13	31-Dec-99	21.34	68.36	
14				
15	Average return, $E(r_{GM})$ and $E(r_{MSFT})$	14.25	62.72	
16	Variance of return, σ^2_{GM} and σ^2_{MSFT}	637.80	1442.92	
17	Standard deviation of return, σ_{GM} and σ_{MSFT}	25.25	37.99	
18	Covariance of returns, $Cov(r_{GM},r_{MSFT})$	-552.10		<-- =COVAR(B4:B13,C4:C13)
19	Correlation of returns, $\rho_{GM,MSFT}$	-0.5755		<-- =CORREL(B4:B13,C4:C13)
20		-0.5755		<-- =B18/(B17*C17)
21				
22	Correlation is symmetric, $\rho_{MSFT,GM}$	-0.5755		<-- =CORREL(C4:C13,B4:B13)

STATISTICAL NOTE: WHY DOES COVARIANCE DEPEND ON THE UNITS OF MEASUREMENT WHEREAS CORRELATION DOESN'T?

Why is the covariance measured in whole numbers 10,000 times bigger than the covariance measured in percentages? Since we've represented percentages as whole numbers, we've essentially multiplied each percentage return by 100. This is how -11.54% becomes -11.54. Since the covariance multiplies the percentages for GM and MSFT together, this means that we've multiplied our previous calculations by $100 * 100 = 10,000$.

The correlation coefficient divides the covariance by the product of the standard deviations,

$$Correlation(r_{GM}, r_{MSFT}) = \frac{Cov(r_{GM}, r_{MSFT})}{\sigma_{GM}\sigma_{MSFT}}$$

In our new calculation, the covariance is 10,000 times bigger, but each standard deviation is 100 times bigger, so that the denominator is also 10,000 times bigger. The result is that the correlation is the same, no matter if the data is measured in percentages or whole numbers.

Fact 2: The correlation between GM and MSFT is the same as the correlation between MSFT and GM. The same holds for the covariance: $Cov(r_{GM}, r_{MSFT}) = Cov(r_{MSFT}, r_{GM})$. The technical jargon for this is that "correlation and covariance are symmetric." To see this in Excel, note that cells B19 (**=Correl(B4:B13,C4:C13)**) and B22 (**=Correl(C4:C13,B4:B13)**) are equal in the above spreadsheet.

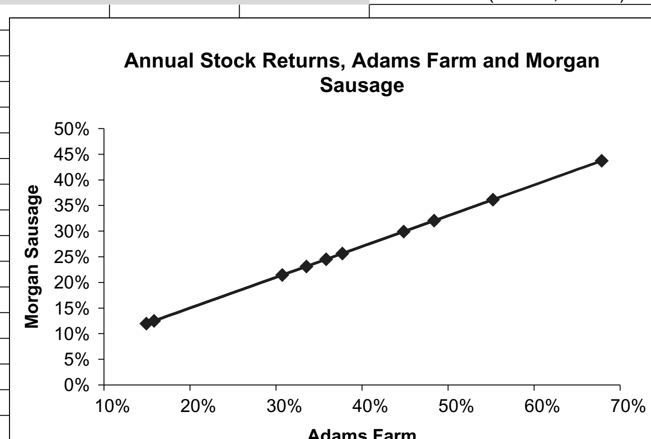
Fact 3: The correlation will always be between +1 and –1. The higher the correlation coefficient is in absolute value, the more the two series move together. If the correlation is either

−1 or +1, then the two series are *perfectly correlated,* which means that knowing one series allows you to predict completely the value of the second series. If the correlation coefficient is between −1 and +1, then the two series move in tandem less than perfectly.

Fact 4: If the correlation coefficient is either +1 or −1, this means that the two returns have a linear relation between them. Since this is not easy to understand, we illustrate with a numerical example. Adams Farm and Morgan Sausage are two shares listed on the Farmers Stock Exchange. For reasons that are difficult to determine, each Morgan Sausage's stock return is equal to 60% of that of Adams Farm plus 3%. We can thus write $r_{Morgan\ Sausage,t} = 3\% + 0.6 * r_{Adams\ Farm,t}$. This means that the return on Morgan Sausage stock is *completely predictable* given the return on Adams Farm stock. Thus, the correlation is either −1 or +1. Since, when Adams Farm's return moves up, so does the return of Morgan Sausage, the correlation is +1.[7]

The Excel spreadsheet that follows confirms that the correlation is +1.

	A	B	C	D
1	**CORRELATION +1** **Adams Farm and Morgan Sausage Stocks** $r_{Morgan\ Sausage,t} = 3\% + 0.6*r_{Adams\ Farm,t}$			
2	**Year**	**Adams Farm stock return**	**Morgan Sausage stock return**	
3	1990	30.73%	21.44%	<-- =3%+0.6*B3
4	1991	55.21%	36.13%	
5	1992	15.82%	12.49%	
6	1993	33.54%	23.12%	
7	1994	14.93%	11.96%	
8	1995	35.84%	24.50%	
9	1996	48.39%	32.03%	
10	1997	37.71%	25.63%	
11	1998	67.85%	43.71%	
12	1999	44.85%	29.91%	
13				
14	Correlation		1.00	<-- =CORREL(B3:B12,C3:C12)



[7]The Farmers Stock Exchange has two other stocks whose returns are related by the equation $r_{Chicken\ Feed,t} = 50\% - 0.8 * r_{Poultry\ Delight,t}$. In this case, the negative coefficient (−0.8) tells us that the correlation between the two sets of returns is −1. (See end-of-chapter Exercise 11.)

0195301501_337-376_ch12.qxd 11/3/05 12:55 PM Page 349

Fact 4 can be written mathematically as follows: Suppose Stock 1 and Stock 2 are *perfectly correlated* (meaning that the correlation is either +1 or −1). Then

$$r_{Stock1,t} = a + b * r_{Stock2,t}, \text{ where } \begin{cases} b > 0 \text{ if the correlation} = +1 \\ b < 0 \text{ if the correlation} = -1 \end{cases}$$

12.4 Portfolio Mean and Variance for a Two-Asset Portfolio

A *portfolio* is a set of stocks or other financial assets. Most people who own stock own more than one stock; they own portfolios of stocks, and the risks they bear relate to the riskiness of their portfolio. In the next chapter we'll start our economic analysis of portfolios. In this section we show you how to compute the mean and variance of a portfolio composed of two stocks. Suppose that between 1990 and 1999 you held a portfolio invested 50% in GM and 50% in MSFT. Column E of the spreadsheet below shows what the annual returns would have been on this portfolio. For example, holding 50% of your portfolio in GM and 50% in Microsoft would have given you a portfolio return of 30.73% in 1990:

$$30.73\% = \underbrace{50\%}_{\substack{\uparrow \\ \text{Proportion of} \\ \text{GM stock in} \\ \text{portfolio}}} * \underbrace{(-11.54\%)}_{\substack{\uparrow \\ \text{Return on} \\ \text{GM stock} \\ \text{in 1990}}} + \underbrace{50\%}_{\substack{\uparrow \\ \text{Proportion of} \\ \text{MSFT stock in} \\ \text{portfolio}}} * \underbrace{72.99\%}_{\substack{\uparrow \\ \text{Return on} \\ \text{MSFT stock} \\ \text{in 1990}}}$$

In cells E17 to E21 we calculate the portfolio return statistics in the same way we calculated the return statistics for the individual assets GM and MSFT.

	A	B	C	D	E	F
1	**CALCULATING PORTFOLIO RETURNS AND THEIR STATISTICS**					
2	Proportion of GM	0.5				
3	Proportion of MSFT	0.5	<-- =1-B2			
4						
5	**Date**	**General Motors GM**	**Microsoft MSFT**		**Portfolio return**	
6	Dec-90	-11.54%	72.99%		30.73%	<-- =$B$2*B6+$B$3*C6
7	Dec-91	-11.35%	121.76%		55.21%	
8	Dec-92	16.54%	15.11%		15.82%	
9	Dec-93	72.64%	-5.56%		33.54%	
10	Dec-94	-21.78%	51.63%		14.93%	
11	Dec-95	28.13%	43.56%		35.84%	
12	Dec-96	8.46%	88.32%		48.39%	
13	Dec-97	19.00%	56.43%		37.71%	
14	Dec-98	21.09%	114.60%		67.85%	
15	Dec-99	21.34%	68.36%		44.85%	
16						
17	Mean	14.25%	62.72%		38.49%	<-- =AVERAGE(E6:E15)
18	Variance	6.38%	14.43%		2.44%	<-- =VARP(E6:E15)
19	St. dev.	25.25%	37.99%		15.62%	<-- =STDEVP(E6:E15)
20	Covariance		-0.0552			
21	Correlation		-0.5755			
22						
23	**Direct calculation of portfolio mean and variance**					
24	Portfolio mean	38.49%	<-- =B2*B17+B3*C17			
25	Portfolio variance	2.44%	<-- =B2^2*B18+B3^2*C18+2*B2*B3*C20			
26	Portfolio st. dev.	15.62%	<-- =SQRT(B25)			

Cells B24 to B26 show that these portfolio statistics can be calculated directly from the statistics for the individual assets. To calculate the portfolio mean using these shortcuts, we first need some notation: Let x_{GM} stand for the proportion of GM stock in the portfolio and let x_{MSFT} denote the proportion of MSFT stock in the portfolio. In our example, $x_{GM} = 0.5$ and $x_{MSFT} = 0.5$ and the portfolio mean return is given by

$$\begin{aligned} \textit{portfolio mean return} &= E(r_p) = x_{GM}E(r_{GM}) + x_{MSFT}E(r_{MSFT}) \\ &= x_{GM}E(r_{GM}) + (1 - x_{GM})E(r_{MSFT}) \end{aligned}$$

Notice the second line of the formula: If we only have two assets in the portfolio, then the proportion of the second asset is "one minus" the proportion of the first asset: $x_{MSFT} = 1 - x_{GM}$.

The formula for the portfolio variance is given by

$$\begin{aligned} \textit{portfolio variance} &= Var(r_p) \\ &= x_{GM}^2 Var(r_{GM}) + x_{MSFT}^2 Var(r_{MSFT}) + 2x_{GM}x_{MSFT}Cov(r_{GM}, r_{MSFT}) \end{aligned}$$

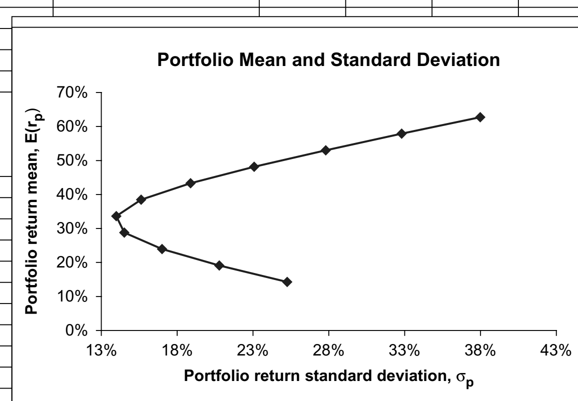
In the spreadsheet below we built a table of the portfolio statistics using the formulas. In the table we vary the proportion of GM stock in the portfolio from 0% to 100% (which means, of course, that the proportion of MSFT stock goes from 100% to 0%).

	A	B	C	D
1	**CALCULATING PORTFOLIO RETURNS AND THEIR STATISTICS FROM THE FORMULAS**			
2		**General Motors GM**	**Microsoft MSFT**	
3	Mean	14.25%	62.72%	
4	Variance	6.38%	14.43%	
5	St. dev.	25.25%	37.99%	
6	Covariance		-5.52%	
7				
8	**Proportion of GM in portfolio**	**Portfolio Variance $Var(r_p)$**	**Portfolio standard deviation σ_p**	**Portfolio mean $E(r_p)$**
9	0%	14.43%	37.99%	62.72%
10	10%	10.76%	32.80%	57.87%
11	20%	7.72%	27.79%	53.03%
12	30%	5.33%	23.08%	48.18%
13	40%	3.56%	18.88%	43.33%
14	50%	2.44%	15.62%	38.49%
15	60%	1.95%	13.98%	33.64%
16	70%	2.11%	14.51%	28.79%
17	80%	2.89%	17.01%	23.95%
18	90%	4.32%	20.78%	19.10%
19	100%	6.38%	25.25%	14.25%

=SQRT(B19)

=A19*$B$3+(1-A19)*$C$3

=A19^2*$B$4+(1-A19)^2*$C$4+2*A19*(1-A19)*$C$6



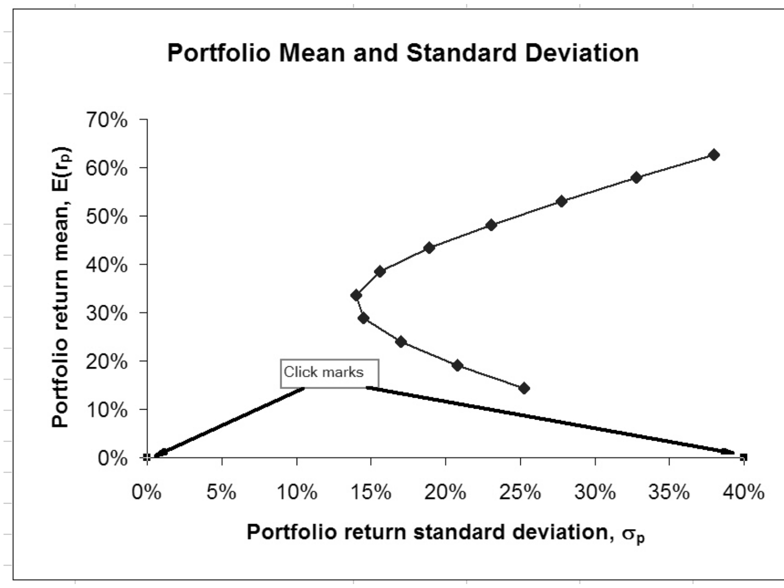
The graph is one that you will see again (lots!) in Chapters 13 and 14. It plots the portfolio standard deviation σ_p on the x-axis and the portfolio mean return $E(r_p)$ on the y-axis. The parabolic shape of the graph is the subject of much discussion in finance, but this is a purely technical chapter—the finance part of the discussion will have to wait until the following chapters.

EXCEL NOTE

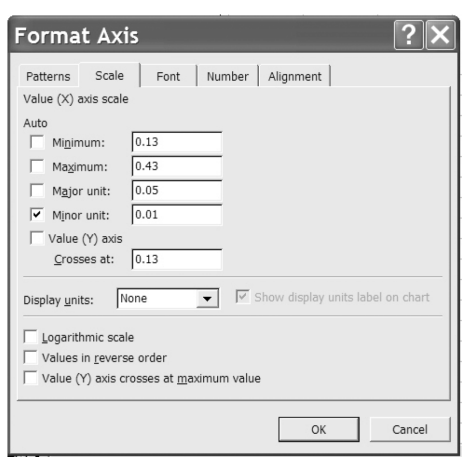
TWO NOTES ABOUT THE GRAPH

Note 1: The graph above is an Excel XY (scatter) plot of the data in the range C9:D19. Notice that we've put the data in a somewhat "unnatural" order: We first compute the variance (cells B9 to B19), then the standard deviation (cells C9 to C19), and only then the expected return (D9 to D19). All this is done to make it easier to use Excel's XY charts, which by default use the leftmost data column as the data for the *x*-axis and data in columns to the right for *y*-axis data. (There are other work-arounds, but they're too cumbersome to explain here.)

Note 2: When we originally made this graph, the *x*-axis went from 0% to 40%.



We "shortened" the *x*-axis by (1) clicking on the axis (you can see the "click marks" on the *x*-axis above), (2) right-clicking the mouse and bringing up the menu for **Format axis**, and (3) changing the **Minimum** and the **Maximum** settings to the following:



12.5 Using Regressions

Linear regression (for short: regression) is a technique for fitting a line to a set of data. Regressions are used in finance to examine the relation between data series. In the chapters that follow we often need to use regressions; we introduce the basic concepts here. We do not discuss the statistical theory behind regressions, but instead we show you how to run a regression and how to use it.

We've divided the discussion into three subsections: First we discuss the mechanics of doing a regression in Excel, then we discuss the meaning of the regression, and finally we discuss alternative ways of doing the regression.

The Mechanics of Doing a Regression in Excel

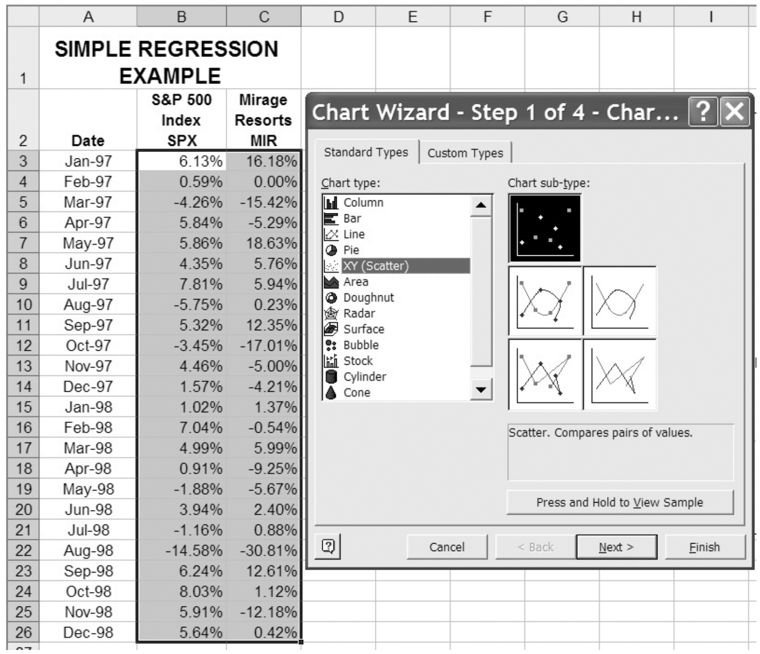
In this subsection we discuss a simple regression example and make little attempt to explain the economic meaning of the regression. Instead, we focus on the mechanics of doing the regression in Excel and leave the economic interpretation for the next subsection.

The table below gives the monthly returns for the S&P 500 Index (stock symbol SPX) and for Mirage Resorts (stock symbol MIR) for 1997 and 1998. The S&P 500 Index includes the 500 largest stocks traded on U.S. stock exchanges, and its performance is roughly indicative of the performance of the U.S. stock market as a whole. We use the regression analysis to see if we can understand the relation between the SPX's returns and MIR's returns—that is, if we can understand the effect of the U.S. stock market on the returns of MIR stock.

Here's the data we examine:

	A	B	C
1	**SIMPLE REGRESSION EXAMPLE**		
2	**Date**	**S&P 500 Index SPX**	**Mirage Resorts MIR**
3	Jan-97	6.13%	16.18%
4	Feb-97	0.59%	0.00%
5	Mar-97	-4.26%	-15.42%
6	Apr-97	5.84%	-5.29%
7	May-97	5.86%	18.63%
8	Jun-97	4.35%	5.76%
9	Jul-97	7.81%	5.94%
10	Aug-97	-5.75%	0.23%
11	Sep-97	5.32%	12.35%
12	Oct-97	-3.45%	-17.01%
13	Nov-97	4.46%	-5.00%
14	Dec-97	1.57%	-4.21%
15	Jan-98	1.02%	1.37%
16	Feb-98	7.04%	-0.54%
17	Mar-98	4.99%	5.99%
18	Apr-98	0.91%	-9.25%
19	May-98	-1.88%	-5.67%
20	Jun-98	3.94%	2.40%
21	Jul-98	-1.16%	0.88%
22	Aug-98	-14.58%	-30.81%
23	Sep-98	6.24%	12.61%
24	Oct-98	8.03%	1.12%
25	Nov-98	5.91%	-12.18%
26	Dec-98	5.64%	0.42%

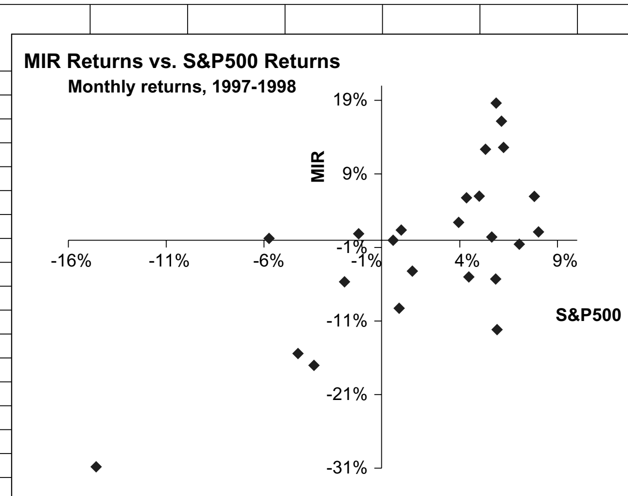
We now use Excel to produce an XY scatter plot of these returns. We use the command **Insert|Chart**, and then the **Chart Wizard** to produce the desired graph:



	A	B	C
1	SIMPLE REGRESSION EXAMPLE		
2	Date	S&P 500 Index SPX	Mirage Resorts MIR
3	Jan-97	6.13%	16.18%
4	Feb-97	0.59%	0.00%
5	Mar-97	-4.26%	-15.42%
6	Apr-97	5.84%	-5.29%
7	May-97	5.86%	18.63%
8	Jun-97	4.35%	5.76%
9	Jul-97	7.81%	5.94%
10	Aug-97	-5.75%	0.23%
11	Sep-97	5.32%	12.35%
12	Oct-97	-3.45%	-17.01%
13	Nov-97	4.46%	-5.00%
14	Dec-97	1.57%	-4.21%
15	Jan-98	1.02%	1.37%
16	Feb-98	7.04%	-0.54%
17	Mar-98	4.99%	5.99%
18	Apr-98	0.91%	-9.25%
19	May-98	-1.88%	-5.67%
20	Jun-98	3.94%	2.40%
21	Jul-98	-1.16%	0.88%
22	Aug-98	-14.58%	-30.81%
23	Sep-98	6.24%	12.61%
24	Oct-98	8.03%	1.12%
25	Nov-98	5.91%	-12.18%
26	Dec-98	5.64%	0.42%

Here's what the chart looks like. As described in Chapter 28 on graphs in Excel, we've gotten rid of the grey background, which is the Excel default.

	A	B	C
1	SIMPLE REGRESSION EXAMPLE		
2	Date	S&P 500 Index SPX	Mirage Resorts MIR
3	Jan-97	6.13%	16.18%
4	Feb-97	0.59%	0.00%
5	Mar-97	-4.26%	-15.42%
6	Apr-97	5.84%	-5.29%
7	May-97	5.86%	18.63%
8	Jun-97	4.35%	5.76%
9	Jul-97	7.81%	5.94%
10	Aug-97	-5.75%	0.23%
11	Sep-97	5.32%	12.35%
12	Oct-97	-3.45%	-17.01%
13	Nov-97	4.46%	-5.00%
14	Dec-97	1.57%	-4.21%
15	Jan-98	1.02%	1.37%
16	Feb-98	7.04%	-0.54%
17	Mar-98	4.99%	5.99%
18	Apr-98	0.91%	-9.25%
19	May-98	-1.88%	-5.67%
20	Jun-98	3.94%	2.40%
21	Jul-98	-1.16%	0.88%
22	Aug-98	-14.58%	-30.81%
23	Sep-98	6.24%	12.61%
24	Oct-98	8.03%	1.12%
25	Nov-98	5.91%	-12.18%
26	Dec-98	5.64%	0.42%

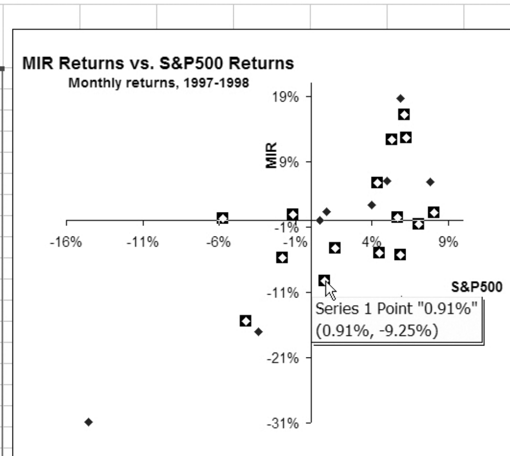


We want to draw a line through the points above, and we want this line to be the "best" line in the sense that it is the closest line you could draw through the points.[8] There are several ways to do this in Excel (as usual . . .). Here's what we do:

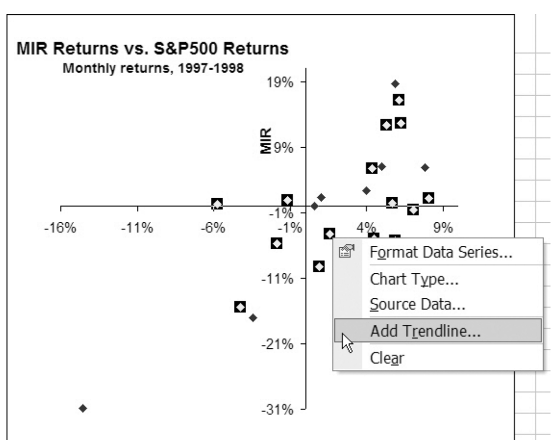
- Click on the points of the graph so that Excel marks all of them. If you have a lot of data points, Excel may mark only some of the points; just ignore this and proceed to the next step. After you do this, you'll see a graph like the one below (notice that Excel shows us the coordinates of the point we happened to point at—in this case the point where the SPX return is 0.91% and the MIR return is −9.25%):

SIMPLE REGRESSION EXAMPLE

	Date	S&P 500 Index SPX	Mirage Resorts MIR
3	Jan-97	6.13%	16.18%
4	Feb-97	0.59%	0.00%
5	Mar-97	-4.26%	-15.42%
6	Apr-97	5.84%	-5.29%
7	May-97	5.86%	18.63%
8	Jun-97	4.35%	5.76%
9	Jul-97	7.81%	5.94%
10	Aug-97	-5.75%	0.23%
11	Sep-97	5.32%	12.35%
12	Oct-97	-3.45%	-17.01%
13	Nov-97	4.46%	-5.00%
14	Dec-97	1.57%	-4.21%
15	Jan-98	1.02%	1.37%
16	Feb-98	7.04%	-0.54%
17	Mar-98	4.99%	5.99%
18	Apr-98	0.91%	-9.25%
19	May-98	-1.88%	-5.67%
20	Jun-98	3.94%	2.40%
21	Jul-98	-1.16%	0.88%
22	Aug-98	-14.58%	-30.81%
23	Sep-98	6.24%	12.61%
24	Oct-98	8.03%	1.12%
25	Nov-98	5.91%	-12.18%
26	Dec-98	5.64%	0.42%

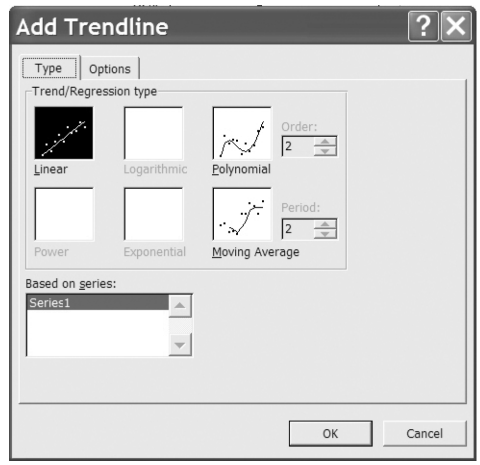


- With the points marked, right-click the mouse and choose **Add Trendline**.

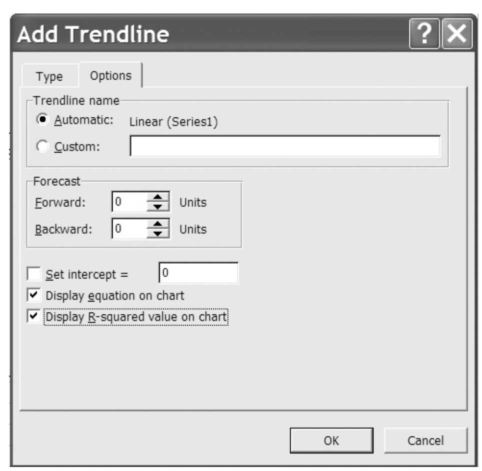


[8] There's a formal statistical definition of "best" and "closest," but we leave that to another course.

- **Add Trendline** brings up the following box, in which we leave the choice **Linear** regression.

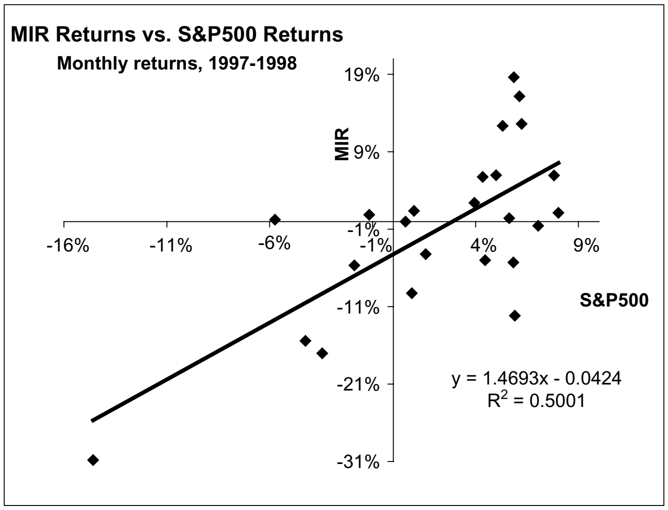


- Before clicking **OK**, we move to the **Options** tab and mark **Display equation on chart** and **Display R-squared value on chart.**



- Now you can click **OK**.

Excel displays the following chart:



The box with the regression results can be moved with the mouse. The text inside the box can be formatted to the font and the size you desire.

What Does the Regression Mean?

The graph above shows the regression line as $y = 1.4693x - 0.0424$, $R^2 = 0.5001$. Since we're trying to understand the effect of the S&P 500 Index on MIR stock, we can attach the following meaning to the variables of the regression line:

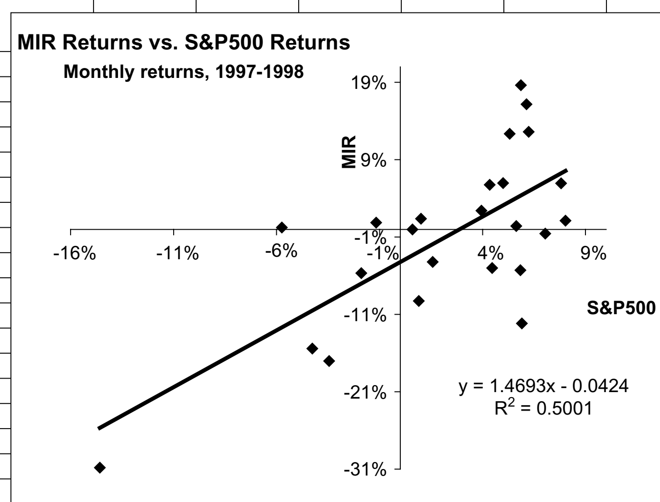
- The "y" of the regression line stands for the monthly percentage return of MIR and the "x" stands for the monthly percentage return of the S&P 500 Index.
- The *slope* of the regression line is 1.4693. This tells us that, on average, a 1% increase in the S&P 500 monthly return caused a 1.4693% increase in the MIR monthly return. Of course, the reverse is true: On average a 1% decrease in the S&P 500 is related to a 1.4693% decrease in MIR's return.
- The fact that the slope of the regression is greater than 1 means that MIR is very sensitive to the S&P 500: Variations (increases or decreases) in the S&P 500 return cause larger variations in the MIR return. We return to this topic in Chapter 14.
- The *intercept* of the regression line is -0.0424. The intercept tells us that in months when the S&P 500 doesn't "move," MIR's return tends to decrease by 4.24%.
- The R^2 (pronounced "r squared") of the regression line says that 50.01% of the variability in the MIR returns is explained by the variability of the S&P 500 returns. This may seem sort of low but it's actually quite respectable: The R^2 of 50% says that half of MIR's return variability is explained by the variability of the S&P 500 Index. The other 50% of the return variability is presumably explained by factors that are unique to MIR. You wouldn't expect much more: If for some strange reason the R^2 were 100%, this would mean that *all* of MIR's returns are explained by the S&P 500 returns, which is clearly nonsense.

The regression line thus allows you to make some interesting predictions about the MIR return based on the S&P 500 return. Suppose you're a financial analyst and you think that this month the S&P 500 Index will go up by 20%. Then based on the regression, you'd expect MIR to increase by $1.4693 * 20\% - 0.0424 = 25.146\%$. Knowing that the R^2 is approximately 50%, only about half of the variability in MIR stock returns is explained by the S&P 500 stock return, and you would thus attach some degree of skepticism to this prediction.

Other Ways of Doing a Regression in Excel

As you might expect, in Excel there are other methods for calculating the slope, intercept, and R^2 of the regression equation. Excel has functions called **Slope**, **Intercept**, **Rsq**. These functions are illustrated below in cells B28, B31, B34. Note that in these functions, the MIR returns come before the S&P 500 returns, so that we write, for example, **Slope(MIR returns,S&P returns)**.

	A	B	C
1	SIMPLE REGRESSION EXAMPLE		
2	Date	S&P 500 Index SPX	Mirage Resorts MIR
3	Jan-97	6.13%	16.18%
4	Feb-97	0.59%	0.00%
5	Mar-97	-4.26%	-15.42%
6	Apr-97	5.84%	-5.29%
7	May-97	5.86%	18.63%
8	Jun-97	4.35%	5.76%
9	Jul-97	7.81%	5.94%
10	Aug-97	-5.75%	0.23%
11	Sep-97	5.32%	12.35%
12	Oct-97	-3.45%	-17.01%
13	Nov-97	4.46%	-5.00%
14	Dec-97	1.57%	-4.21%
15	Jan-98	1.02%	1.37%
16	Feb-98	7.04%	-0.54%
17	Mar-98	4.99%	5.99%
18	Apr-98	0.91%	-9.25%
19	May-98	-1.88%	-5.67%
20	Jun-98	3.94%	2.40%
21	Jul-98	-1.16%	0.88%
22	Aug-98	-14.58%	-30.81%
23	Sep-98	6.24%	12.61%
24	Oct-98	8.03%	1.12%
25	Nov-98	5.91%	-12.18%
26	Dec-98	5.64%	0.42%
27			
28	Slope	1.4693	<-- =SLOPE(C3:C26,B3:B26)
29		1.4693	<-- =COVAR(C3:C26,B3:B26)/VARP(B3:B26)
30			
31	Intercept	-0.0424	<-- =INTERCEPT(C3:C26,B3:B26)
32		-0.0424	<-- =AVERAGE(C3:C26)-B28*AVERAGE(B3:B26)
33			
34	R-squared	0.5001	<-- =RSQ(C3:C26,B3:B26)
35		0.5001	<-- =CORREL(C3:C26,B3:B26)^2



The slope, intercept, and R^2 can also be calculated using **Average**, **Covar**, **Varp**, and **Correl** (cells B29, B32, B35). Look at the alternative definitions of each of the regression variables:

- The regression slope can be computed with the **Slope** function (cell B28), but as shown in cell B29 it is also equal to $Covariance(S\&P, MIR)/Var(S\&P)$.
- The regression intercept can be computed with the **Intercept** function (cell 31), but as shown in cell B32 it is also equal to $Average(MIR) - slope * Average(S\&P)$.
- The regression R^2 can be computed with the **Rsq** function (cell B34), but as shown in cell B35 it is also equal to the squared correlation between the S&P and MIR: $[Correlation(S\&P, MIR)]^2$.

12.6 Portfolio Statistics for Multiple Assets (Advanced Topic)

This section discusses a slightly more advanced topic, which is used only in Appendix 13.2 of Chapter 13. You can skip it on first reading. In Section 12.4, we discussed the calculation of the portfolio mean and variance for a two-asset portfolio. In this section we discuss the calculation for a portfolio composed of more than two assets.

In order to set the scene, we introduce some notation. Suppose that we have *three* stocks, and that for each stock i ($i = 1, 2, 3$) we have computed the mean $E(r_i)$ and the variance $\sigma_i^2 = Var(r_i)$ of the stock's returns. Furthermore, suppose that for each pair of stocks i and j, we have calculated the covariance of the returns $Cov(r_i, r_j)$. Here's an example:

	A	B	C	D	E
1	**PORTFOLIO RETURNS FOR A THREE-STOCK PORTFOLIO**				
2	**Year ending**	**General Motors GM**	**Microsoft MSFT**	**Heinz HNZ**	
3	Dec-90	-11.54%	72.99%	2.46%	
4	Dec-91	-11.35%	121.76%	14.54%	
5	Dec-92	16.54%	15.11%	16.89%	
6	Dec-93	72.64%	-5.56%	-15.95%	
7	Dec-94	-21.78%	51.63%	6.55%	
8	Dec-95	28.13%	43.56%	39.81%	
9	Dec-96	8.46%	88.32%	11.56%	
10	Dec-97	19.00%	56.43%	45.89%	
11	Dec-98	21.09%	114.60%	14.11%	
12	Dec-99	21.34%	68.36%	-27.44%	
13					
14	Average	14.25%	62.72%	10.84%	<-- =AVERAGE(D3:D12)
15	Variance	0.0638	0.1443	0.0440	<-- =VARP(D3:D12)
16	Sigma	25.25%	37.99%	20.98%	<-- =STDEVP(D3:D12)
17					
18	**Covariances**				
19	$Cov(r_{GM},r_{MSFT})$	-0.0552	<-- =COVAR(B3:B12,C3:C12)		
20	$Cov(r_{GM},r_{HNZ})$	-0.0096	<-- =COVAR(B3:B12,D3:D12)		
21	$Cov(r_{MSFT},r_{HNZ})$	0.0092	<-- =COVAR(C3:C12,D3:D12)		

Now suppose we form a portfolio composed of the following proportions of each of the stocks: $x_{GM} = 20\%$, $x_{MSFT} = 50\%$, $x_{HNZ} = 1 - x_{GM} - x_{MSFT} = 30\%$. Cells G3 to G12 in the spreadsheet below show you the returns of this portfolio, and cells G14 to G16 compute the portfolio's mean return, variance, and standard deviation:

	A	B	C	D	E	F	G	H
1	**PORTFOLIO RETURNS FOR A THREE-STOCK PORTFOLIO**							
2	**Year ending**	**General Motors GM**	**Microsoft MSFT**	**Heinz HNZ**			**Portfolio return**	
3	Dec-90	-11.54%	72.99%	2.46%			34.92%	<-- =0.2*B3+0.5*C3+0.3*D3
4	Dec-91	-11.35%	121.76%	14.54%			62.97%	<-- =0.2*B4+0.5*C4+0.3*D4
5	Dec-92	16.54%	15.11%	16.89%			15.93%	
6	Dec-93	72.64%	-5.56%	-15.95%			6.96%	
7	Dec-94	-21.78%	51.63%	6.55%			23.42%	
8	Dec-95	28.13%	43.56%	39.81%			39.35%	
9	Dec-96	8.46%	88.32%	11.56%			49.32%	
10	Dec-97	19.00%	56.43%	45.89%			45.78%	
11	Dec-98	21.09%	114.60%	14.11%			65.75%	
12	Dec-99	21.34%	68.36%	-27.44%			30.22%	
13								
14	Average	14.25%	62.72%	10.84%	<-- =AVERAGE(D3:D12)	Average	37.46%	<-- =AVERAGE(G3:G12)
15	Variance	0.0638	0.1443	0.0440	<-- =VARP(D3:D12)	Variance	0.0331	<-- =VARP(G3:G12)
16	Sigma	25.25%	37.99%	20.98%	<-- =STDEVP(D3:D12)	Sigma	18.21%	<-- =STDEVP(G3:G12)
17								
18	**Covariances**						**Alternative calculation of portfolio statistics**	
19	$Cov(r_{GM}, r_{MSFT})$	-0.0552	<-- =COVAR(B3:B12,C3:C12)			**Average**	37.46%	<-- =0.2*B14+0.5*C14+0.3*D14
20	$Cov(r_{GM}, r_{HNZ})$	-0.0096	<-- =COVAR(B3:B12,D3:D12)			**Variance**	0.0331	<-- =0.2^2*B15+0.5^2*C15+0.3^2*D15 +2*0.2*0.5*B19+2*0.2*0.3*B20+2*0.5*0.3*B21
21	$Cov(r_{MSFT}, r_{HNZ})$	0.0092	<-- =COVAR(C3:C12,D3:D12)			**Sigma**	18.21%	<-- =SQRT(G20)

If you look at cells G19 to G21, you'll see that there is a more straightforward way of doing the same calculations, based on the following formulas:

$$\begin{aligned}
\textit{average portfolio return} \text{ (cell G19)} &= E(r_p) = x_{GM}E(r_{GM}) + x_{MSFT}E(r_{MSFT}) + x_{HNZ}E(r_{HNZ}) \\
\textit{portfolio variance} \text{ (cell G20)} &= Var(r_p) = x_{GM}^2 Var(r_{GM}) + x_{MSFT}^2 Var(r_{MSFT}) + x_{HNZ}^2 Var(r_{HNZ}) \\
&\quad + 2x_{GM}x_{MSFT}Cov(r_{GM}, r_{MSFT}) + 2x_{GM}x_{HNZ}Cov(r_{GM}, r_{HNZ}) \\
&\quad + 2x_{MSFT}x_{HNZ}Cov(r_{MSFT}, r_{HNZ}) \\
\textit{portfolio standard deviation} \text{ (cell G21)} &= \sqrt{\textit{portfolio variance}\text{(cell G20)}}
\end{aligned}$$

These formulas generalize to any number of assets: If we have a portfolio composed of N assets, and we know all the expected returns, variances, and covariances, then:

- The portfolio's expected return is the weighted average of the individual asset returns. Denoting the portfolio weights by $\{x_1, x_2, \ldots, x_N\}$, the portfolio expected return is

$$\begin{aligned}
E(r_p) &= x_1 E(r_1) + x_2 E(r_2) + \cdots + x_N E(r_N) \\
&= \sum_{i=1}^{N} x_i E(r_i)
\end{aligned}$$

- The portfolio's variance of return is the sum of the following two expressions:
  - The sum of each asset's variance, weighted by the *square* of the asset's portfolio proportion: $x_1^2 Var(r_1) + x_2^2 Var(r_2) + \cdots + x_N^2 Var(r_N)$.

- The sum of twice each of the covariances, weighted by the *product* of the asset proportions:

$$2x_1x_2Cov(r_1, r_2) + 2x_1x_3Cov(r_1, r_3) + \cdots + 2x_1x_NCov(r_1, r_N) + 2x_2x_3Cov(r_2, r_3) + \cdots + 2x_2x_NCov(r_2, r_N) + \cdots + 2x_{N-1}x_NCov(r_{N-1}, r_N)$$

CONCLUSION

Information about stocks—their prices, dividends, and returns—produces mounds of data. Statistics is a way of dealing with these large masses of data. This chapter has given you the necessary statistical techniques to do typical finance computations related to stocks. We've shown how to compute stock returns from basic data about stock prices, dividends, and stock splits. We've also shown how to compute the mean return (also called the average return), the variance and standard deviation of returns, and the covariance between the returns of two different stocks.

Stocks are most often combined into portfolios, and this chapter has shown you how to compute the mean and standard deviation of a portfolio's return. It also introduced you to regression analysis, which allows you to relate the returns of two stocks to each other.

In succeeding chapters we will use these statistical techniques to do financial analysis of individual stocks and stock portfolios.

EXERCISES

1. Here is the stock price history of HighTech Corp. and LowTech Corp.

	A	B	C
1		**HighTech Corp. Stock price**	**LowTech Corp. Stock price**
2	31-Dec-91	75.00	40.00
3	31-Dec-92	86.25	45.20
4	31-Dec-93	125.32	55.60
5	31-Dec-94	91.64	48.37
6	31-Dec-95	100.80	32.88
7	31-Dec-96	145.93	61.64
8	31-Dec-97	151.21	75.82
9	31-Dec-98	196.57	97.05
10	31-Dec-99	226.05	109.66
11	31-Dec-00	89.00	122.99

Calculate the following:

(a) The annual returns for each stock.

(b) The mean (average) return for the ten-year period for each firm. Which stock has the higher average return?

(c) The variance and the standard deviation of returns, for the ten-year period for each firm. Which stock is riskier?

Note: Data for these exercises can be found on the disk that accompanies this book.

(d) The covariance and correlation of the returns for each firm. Use two formulas to compute the correlation: The Excel formula **Correl** and the definition

$$Correlation(r_A, r_B) = \frac{Cov(r_A, r_B)}{\sigma_A \sigma_B}$$

(e) If you had to choose between the two stocks, which would you choose? Explain briefly.

2. Below are price data for three mutual funds:

	A	B	C	D
1	**DATA ON THREE MUTUAL FUNDS**			
2	**Date**	**Scudder Development Fund**	**Value Line Leveraged Growth Fund**	**Fidelity Fund**
3	4-Jan-93	24.34	17.47	9.47
4	3-Jan-94	24.2	20.32	11.39
5	3-Jan-95	20.87	19.15	11.19
6	2-Jan-96	30.35	24.45	15.25
7	2-Jan-97	30.94	26.95	18.46
8	2-Jan-98	31.28	32.08	23.44
9	4-Jan-99	33.32	47.19	31.04
10	3-Jan-00	36.06	49.12	35.36
11	2-Jan-01	33.89	47.23	33.82
12	2-Jan-02	20.01	37.31	28.46
13	2-Jan-03	13.79	26.87	21.55

(a) Compute the annual returns on the funds for the period.

(b) Compute the mean, variance, and standard deviation of the fund returns.

(c) Graph the fund returns and the dates.

(d) Calculate the correlations of the fund returns.

(e) If the historical information correctly predicts future returns (is this reasonable?), which fund would you choose?

3. Here are the monthly stock price data for Ford Corporation and GM Corporation:

	A	B	C	D
1	**PRICES FOR FORD AND GM STOCK**			
2	**Date**	**Ford**	**GM**	
3	8-Nov-99	24.44	66.08	
4	1-Dec-99	25.79	65.09	
5	3-Jan-00	24.32	72.14	
6	1-Feb-00	20.35	68.54	
7	1-Mar-00	22.45	74.63	
8	3-Apr-00	27.00	84.37	
9	1-May-00	23.95	64.02	
10	1-Jun-00	22.08	52.63	
11	3-Jul-00	24.17	51.61	
12	1-Aug-00	21.95	63.97	
13	1-Sep-00	23.14	59.40	
14	2-Oct-00	23.98	56.77	
15	1-Nov-00	20.89	45.64	
16	1-Dec-00	21.52	46.96	
17	2-Jan-01	26.16	49.51	
18	1-Feb-01	25.30	51.77	

Calculate the following:

(a) Monthly returns for each firm.

(b) Covariance between returns of Ford Corporation and GM Corporation.

(c) Correlation between returns of Ford Corporation and GM Corporation.

4. By using the returns of Ford Corp. and GM Corp. that you calculated in Exercise 3, perform a regression of Ford's returns versus GM's returns. Report:
   - The slope of the regression.
   - The value of the intercept.
   - The R^2 of the regression.

   Is the mutual impact of the two company's sales (one on the other) large or small? Explain.

5. Here are the stock price and dividend data for Kellogg Co.:

	A	B	C
1	**KELLOGG PRICE AND DIVIDEND DATA**		
2		**Price**	**Dividend during year**
3	31-Dec-89	64.62	
4	31-Dec-90	78.00	1.44
5	31-Dec-91	56.75	2.15
6	31-Dec-92	62.12	1.16
7	31-Dec-93	53.75	1.32
8	31-Dec-94	55.00	1.40
9	31-Dec-95	76.62	1.50
10	31-Dec-96	69.62	1.62
11	31-Dec-97	46.38	1.28
12	31-Dec-98	40.62	0.90
13	31-Dec-99	24.25	0.98
14	31-Dec-00	26.20	1.00
15	31-Dec-01	30.86	1.00
16	31-Dec-02	33.40	1.00

   (a) Calculate the dividend-adjusted returns for each of the years, their mean, and their standard deviation.

   (b) Stock analysts like to talk about the *dividend yield*—the dividend divided by the stock price. Compute the annual dividend yield for Kellogg (define it as (*dividends over the year*)/(*stock price at begining of year*)) and compute its statistics (mean and standard deviation) over the period.

   (c) If you bought Kellogg stock and had no intention of ever selling it, why might you be interested in the stock's dividend yield?

6. Below are the stock price, dividend, and split data for IBM. Calculate the dividend and split-adjusted returns for each of the years, their mean, and their standard deviation.

	A	B	C	D
1	**IBM PRICE, DIVIDEND AND SPLIT DATA**			
2		**Closing price**	**Dividend during year**	**Other information**
3	31-Dec-89	98.62		
4	31-Dec-90	126.75		
5	31-Dec-91	90.00		
6	31-Dec-92	51.50		
7	31-Dec-93	56.50		
8	31-Dec-94	72.12		
9	31-Dec-95	108.50		
10	31-Dec-96	156.88		
11	31-Dec-97	98.75		2 for 1 split (May 97)
12	31-Dec-98	183.25		
13	31-Dec-99	112.25		2 for 1 split (May 99)
14	31-Dec-00	112.00		
15	31-Dec-01	107.89		
16	31-Dec-02	78.20	0.30	

7. Compute the covariance and correlation coefficient between Kellogg and IBM (Exercises 5 and 6). Are there any advantages to diversifying between Kellogg and IBM?
8. Here are the stock price and split data for HeavySteel Corporation.

	A	B	C
1	**HEAVYSTEEL CORPORATION**		
2		**Closing stock price**	**Stock splits**
3	31-Dec-90	11.24	
4	31-Dec-91	11.98	
5	31-Dec-92	10.23	
6	31-Dec-93	11.02	2 for 1
7	31-Dec-94	12.56	
8	31-Dec-95	13.45	
9	31-Dec-96	15.36	1.5 for 1
10	31-Dec-97	16.01	
11	31-Dec-98	17.23	
12	31-Dec-99	15.23	

(a) Calculate the split-adjusted returns for each year and its statistics (mean and standard deviation).

(b) If you bought 100 shares of this stock in the beginning of 1990 and during the period of ten years never sold or bought additional shares, how many shares would you have by the end of 2000?

9. A *reverse split* is just like a split, but only in a reverse direction. For example, in a 1 for 2 reverse split, you receive 1 share for every 2 shares you hold. How would your answers to the previous question change if you learned that in 1999 the firm did a 3 for 4 reverse split?

10. Consider two companies: Young Corporation and Mature Corporation. Young Corporation grows very rapidly, does not pay any dividends, and retains all its profits. Mature Corporation stopped growing a long time ago, generates sizable cash flows, and pays out dividends.

	A	B	C	D
1		**Young Corp.**	**Mature Corp.**	
2		Share price	Share price	Dividend per share
3	31-Dec-90	32.56	78.50	0.00
4	31-Dec-91	34.50	82.50	0.00
5	31-Dec-92	38.98	84.50	1.00
6	31-Dec-93	44.50	81.60	0.00
7	31-Dec-94	40.20	79.60	1.50
8	31-Dec-95	39.50	80.96	1.50
9	31-Dec-96	38.45	82.65	0.00
10	31-Dec-97	37.50	83.69	2.00
11	31-Dec-98	43.58	82.79	2.00
12	31-Dec-99	50.30	81.97	0.00

(a) Calculate Young's yearly returns.

(b) Calculate Mature's yearly returns.

(c) Which is the better investment of the two? Give a brief explanation.

11. Chicken Feed and Poultry Delight are two stocks traded on the Farmers Stock Exchange. A statistician has determined that the returns on the two stocks are related by the equation $r_{Chicken\ Feed,t} = 50\% - 0.8 * r_{Poultry\ Delight,t}$. Show that the correlation between the two sets of returns is -1. Use the following template:

	A	B	C
2	**Year**	**Poultry Delight stock return**	**Chicken Feed stock return**
3	1990	30.73%	
4	1991	55.21%	
5	1992	15.82%	
6	1993	33.54%	
7	1994	14.93%	
8	1995	35.84%	
9	1996	48.39%	
10	1997	37.71%	
11	1998	67.85%	
12	1999	44.85%	
13			
14	Correlation		

12. Below are the annual returns of two assets. Fill in the blanks and graph the returns of the portfolios (rows 13–27).

	A	B	C
1		**Asset 1**	**Asset 2**
2	31-Dec-90	12.56%	7.56%
3	31-Dec-91	13.50%	8.56%
4	31-Dec-92	14.23%	4.56%
5	31-Dec-93	15.23%	2.12%
6	31-Dec-94	14.23%	1.23%
7	31-Dec-95	12.23%	0.26%
8	31-Dec-96	10.23%	3.25%
9	31-Dec-97	5.26%	4.89%
10	31-Dec-98	4.25%	5.56%
11	31-Dec-99	2.23%	6.45%
12			
13	Average return		
14	Return variance		
15	Covariance		
16	**Proportion of asset 1**	**Portfolio standard deviation**	**Portfolio mean return**
17	0		
18	0.1		
19	0.2		
20	0.3		
21	0.4		
22	0.5		
23	0.6		
24	0.7		
25	0.8		
26	0.9		
27	1		

13. Here are data on the stock prices and returns of General Electric, Boeing, and S&P 500 Index.

	A	B	C	D	E	F	G
1	MONTHLY RETURNS ON GE, BOEING, S&P500, 2000						
2	Date	GE	GE Return	Boeing	Boeing return	S&P500	S&P return
3	Jan-02	37.15		40.22		1130.20	
4	Feb-02	38.50	3.63%	45.33	12.71%	1106.73	-2.08%
5	Mar-02	37.40	-2.86%	47.59	4.99%	1147.39	3.67%
6	Apr-02	31.55	-15.64%	43.99	-7.56%	1076.92	-6.14%
7	May-02	31.14	-1.30%	42.23	-4.00%	1067.14	-0.91%
8	Jun-02	29.05	-6.71%	44.55	5.49%	989.82	-7.25%
9	Jul-02	32.20	10.84%	41.11	-7.72%	911.62	-7.90%
10	Aug-02	30.15	-6.37%	36.87	-10.31%	916.07	0.49%
11	Sep-02	24.65	-18.24%	33.95	-7.92%	815.28	-11.00%
12	Oct-02	25.25	2.43%	29.59	-12.84%	885.76	8.64%
13	Nov-02	27.12	7.41%	34.05	15.07%	936.31	5.71%
14	Dec-02	24.35	-10.21%	32.99	-3.11%	879.82	-6.03%
15							
16	Average return						
17	Standard deviation						
18	Covariances						
19	Cov(GE,Boeing)						
20	Cov(GE,SP)						
21	Cov(Boeing,SP)						
22							
23	Correlations						
24	Correlation(GE,Boeing)						
25	Correlation(GE,SP)						
26	Correlation(Boeing,SP)						
27							
28	Portfolio proportions						
29	GE	0.5					
30	Boeing	0.3					
31	S&P	0.2	<-- =1-B30-B29				
32							
33	Portfolio return						
34	Portfolio standard deviation						

Calculate the highlighted cells.

14. Go to http://finance.yahoo.com. Download monthly adjusted stock price data for Oracle Corporation (ORCL), Microsoft Corporation (MSFT), Dell Corporation (DELL), and Gateway Corporation (GTW) for 1998 and 1999. Also, download the same data for S&P 500 Index (SPX) for the same period.[9] Answer the following questions:
    (a) What is the mean return, variance, and standard deviation of a portfolio consisting of the four stocks, where wealth is allocated equally among each stock?
    (b) On average, would you be better off by investing in this portfolio or by investing in S&P 500 Index, during the period of two years?
    (c) What is the sensitivity of your portfolio to the movements of S&P 500 Index? You will have to perform a regression of the portfolio returns versus S&P 500 returns and report the results.

[9]Recall that when you download data from Yahoo into Excel, it is already adjusted for stock splits and dividends.

15. By using information provided in Exercise 14, perform a regression of the portfolio returns versus. S&P 500 Index returns for the period of 24 months. Report the slope of the regression, its intercept, and R^2. Explain what each of these numbers tell you.
16. (This is a hard question!) On the disk that comes with this book, you will find two years of monthly unadjusted and adjusted stock price data for AT&T Corporation (symbol: T).
    (a) Calculate the cumulative adjustment factor for AT&T stock.
    (b) What two interesting things happened in November 2002 and what happened to the cumulative adjustment factor in this month? Can you explain?

Here's the data:

	A	B	C	D	E	F	G	H	I
1	**Date**	**Open**	**High**	**Low**	**Close**	**Volume**	**Adj. Close***	**Cumulative Adjustment factor**	
2	Dec 02			$0.19 Cash Dividend					
3	Dec 02	28.54	28.88	25.11	**26.11**	4,932,428	26.11		
4	Nov 02			$8.48 Cash Dividend					
5	Nov 02			1:5 Stock Split					
6	Nov 02	12.94	28.25	12.84	**28.04**	13,146,915	28.04		<-- AT&T Spins Off AT&T Broadband To Shareowners And Completes AT&T Broadband Merger With Comcast
7	Oct 02	12.1	13.64	10.45	**13.04**	14,453,869	65.2		1 to 5 Reverse Split
8	Sep 02			$0.04 Cash Dividend					
9	Sep 02	11.95	13.79	11.2	**12.01**	15,095,745	60.05		
10	Aug 02	10.12	12.85	8.69	**12.22**	17,147,918	61.1		
11	Jul 02	10.5	10.55	8.2	**10.18**	18,639,136	50.9		
12	Jun 02			$0.04 Cash Dividend					
13	Jun 02	11.85	12.4	9.09	**10.7**	29,520,930	53.5		
14	May 02	13.2	14.3	11.76	**11.97**	17,814,400	59.85		
15	Apr 02	15.74	15.85	12.66	**13.12**	15,936,609	65.6		
16	Mar 02			$0.04 Cash Dividend					
17	Mar 02	15.8	16.48	15	**15.7**	11,042,700	78.5		
18	Feb 02	17.55	17.91	14.18	**15.54**	16,401,442	77.7		
19	Jan 02	18.48	19.25	16.65	**17.7**	11,919,185	88.5		
20	Dec 01			$0.04 Cash Dividend					
21	Dec 01	17.35	18.75	15.8	**18.14**	14,846,490	90.7		
22	Nov 01	15.33	17.85	14.75	**17.49**	10,987,857	87.45		
23	Oct 01	19.15	20	15.17	**15.25**	15,015,643	76.25		
24	Sep 01			$0.04 Cash Dividend					
25	Sep 01	19.01	19.64	16.5	**19.3**	15,798,733	96.5		
26	Aug 01	20.32	20.95	18.66	**19.04**	7,457,491	95.2		
27	Jul 01			$5.52 Cash Dividend					
28	Jul 01	21.75	23	18.1	**20.21**	16,556,647	101.05		
29	Jun 01			$0.04 Cash Dividend					
30	Jun 01	21.16	22.16	19.82	**22**	11,332,052	110		
31	May 01	22.58	23.1	20.48	**21.17**	15,562,513	105.85		
32	Apr 01	21.3	23.27	19.85	**22.28**	12,075,000	111.4		
33	Mar 01			$0.04 Cash Dividend					
34	Mar 01	22.8	24.6	20.6	**21.3**	12,662,459	106.5		
35	Feb 01	23.95	24.53	20.2	**23**	12,220,989	115		
36	Jan 01	17.37	25.15	17.25	**23.99**	20,407,609	119.95		
37	Dec 00			$0.04 Cash Dividend					
38	Dec 00	19.44	22.69	16.5	**17.25**	23,385,210	86.25		
39	Nov 00	22.62	22.94	18.25	**19.62**	20,863,095	98.1		
40	Oct 00	29	30	21.25	**23.19**	24,254,945	115.95		
41	Sep 00			$0.22 Cash Dividend					
42	Sep 00	31.62	32.94	27.25	**29**	19,280,690	145		
43	Aug 00	30.94	32.94	29.62	**31.62**	17,828,760	158.1		
44	Jul 00	31.81	35.19	30.5	**30.94**	19,562,070	154.7		
45	Jun 00			$0.22 Cash Dividend					
46	Jun 00	34.94	37.75	31.25	**31.81**	20,312,436	159.05		
47	May 00	46.31	49	33.63	**34.94**	25,649,081	174.7		
48	Apr 00	56.69	58.81	45.88	**45.88**	12,616,194	229.4		
49	Mar 00			$0.22 Cash Dividend					
50	Mar 00	49.38	61	47.5	**56.31**	13,692,547	281.55		
51	Feb 00	52.75	53	44.31	**49.38**	10,648,485	246.9		
52	Jan 00	50.81	56	47.5	**52.75**	11,964,045	263.75		
53	Dec 99			$0.22 Cash Dividend					
54	Dec 99	55.88	58.69	49.88	**50.81**	9,812,559	254.05		
55	Nov 99	47.13	61	44.94	**55.88**	13,277,338	279.4		
56	Oct 99	43.5	49.06	41.5	**46.75**	11,850,266	233.75		
57	Sep 99			$0.22 Cash Dividend					
58	Sep 99	45.38	48.81	41.81	**43.5**	10,775,514	217.5		
59	Aug 99	52.13	52.81	44.25	**45**	12,892,813	225		
60	Jul 99	55.94	59	51.75	**52.13**	9,257,600	260.65		
61	Jun 99			$0.22 Cash Dividend					
62	Jun 99	55.5	56.88	52.38	**55.81**	10,673,172	279.05		
63	May 99	51	63	50.88	**55.5**	14,542,265	277.5		
64	Apr 99			3:2 Stock Split					
65	Apr 99	79.81	89.5	50.06	**50.5**	13,690,428	252.5		
66	Mar 99			$0.33 Cash Dividend					
67	Mar 99	82.12	89	75.87	**79.81**	9,906,500	266.03		
68	Feb 99	91.94	95.12	82.12	**82.12**	8,755,210	273.73		
69	Jan 99	76.5	96.12	76.5	**90.75**	10,024,863	302.5		

17. Explain why each of the following statements is correct or incorrect:
    (a) Diversification reduces risk because prices of stocks do not usually move exactly together.
    (b) The expected return on a portfolio is a weighted average of the expected returns on the individual securities.
    (c) The standard deviation of returns on a portfolio is equal to the weighted average of the standard deviations on the individual securities if these returns are completely uncorrelated.
18. Suppose that the annual returns on two stocks (A and B) are perfectly negatively correlated, and that $r_A = 0.05$, $r_B = 0.15$, $\sigma_A = 0.1$, $\sigma_B = 0.4$. Assuming that there are no arbitrage opportunities, what must the one-year interest rate be?
19. Assume that an individual can either invest all of her resources in one of two securities A or B; or alternatively, she can diversify her investment between the two. The distribution of the returns are as follows:

	A	B	C	D
1	Security A		Security B	
2	Return	Probability	Return	Probability
3	-10%	0.5	-20%	0.5
4	50%	0.5	60%	0.5

Assume that the correlation between the returns from the two securities is zero.
    (a) Calculate each security's expected return, variance, and standard deviation.
    (b) Calculate the probability distribution of the returns on a *mixed portfolio* comprised of equal proportions of securities A and B. Also calculate the expected return, variance, and standard deviation.
    (c) Calculate the expected return and the variance of a mixed portfolio comprised of 75% of security A and 25% of security B.
20. The correlations between the returns of three stocks A, B, and C are given in the following table:

	A	B	C	D
1	**Stock**	A	B	C
2	A	1.00	0.80	0.10
3	B		1.00	0.15
4	C			1.00

The expected rates of return on A, B, and C are 16%, 12%, and 15%, respectively. The corresponding standard deviations of the returns are 25%, 22%, and 25%.
    (a) What is the standard deviation of a portfolio invested 25% in stock A, 25% in stock B, and 50% in stock C?
    (b) You plan to invest 50% of your money in the portfolio constructed in part (a) of this exercise and 50% in the risk-free asset. The risk-free interest rate is 5%. What is the expected return on this investment? What is the standard deviation of the return on this investment?
21. You believe that there is a 15% chance that stock A will decline by 10% and an 85% chance that it will increase by 15%. Correspondingly, there is a 30% chance that stock B will decline by 18% and a 70% chance that it will increase by 22%. The correlation coefficient between the two stocks is 0.55. Calculate the expected return, the variance, and the standard deviation for each stock. Then calculate the covariance between their returns.

22. Outdoorsy people know that the crickets chirp faster when the temperature is warmer. Some evidence for this can be found in a book published in 1948 by Harvard physics professor George W. Pierce.[10] Pierce's book includes the table below, which relates the average number of cricket chirps per minute to the temperature at which the data was recorded. Plot the data in an Excel graph and use regression to determine the (approximate) relation between the number of chirps per second and the temperature. If you detect 19 chirps per second, what would you guess the temperature to be? What about 22 chirps per second? (We know this problem has nothing to do with finance, but it's interesting!)

	A	B
4	Chirps per second	Temperature in Farenheit
5	20.0	88.60
6	16.0	71.60
7	19.8	93.30
8	18.4	84.30
9	17.1	80.60
10	15.5	75.20
11	14.7	69.70
12	17.1	82.00
13	15.4	69.40
14	16.2	83.30
15	15.0	79.60
16	17.2	82.60
17	16.0	80.60
18	17.0	83.50
19	14.4	76.30

23. Economists have long believed that the more money printed, the higher will be long-term interest rates. Evidence for this view can be found in the table below, which gives long-term government bond rates for 31 countries and the corresponding growth rate of money supply for each country.[11]
    (a) Plot the data and use a regression to find the relation between the money growth and the long-term bond interest rate.
    (b) If a country has zero money growth, what is its predicted long-term bond interest rate?
    (c) The monetary authorities in your country are considering increasing the money growth rate by 1% from its current level. Predict by how much this will increase the long-term bond interest rate.
    (d) Do you find the evidence in the table convincing? (Discuss briefly the R^2 of the regression.)

---

[10]Additional facts: Cricket chirping is produced by the rapid sliding of the cricket's wings one over the other. The higher the temperature, the faster the crickets slide their wings. George W. Pierce's book is called *The Songs of Insects* and was published by Harvard University Press.

[11]The data was first presented in an article entitled "Money and Interest Rates," by Cyril Monnet and Warren Weber in the *Federal Reserve Bank of Minneapolis Quarterly Review,* Fall 2001. My thanks to the authors for providing me with an Excel version of their data.

	A	B	C	D	E	F	G
38	**MONEY GROWTH AND BOND INTEREST RATES**						
39	**Country**	**Average mone growth**	**Average long-term bond interest rate**		**Country**	**Average money growth**	**Average long-term bond interest rate**
40	US	5.65%	7.40%		New Zealand	10.29%	8.81%
41	Austria	6.82%	7.80%		South Africa	14.14%	11.11%
42	Belgium	5.20%	8.22%		Honduras	16.20%	15.57%
43	Denmark	9.43%	10.36%		Jamaica	19.88%	15.35%
44	France	8.15%	8.49%		Netherlands Antilles	4.36%	9.40%
45	Germany	8.00%	7.20%		Trinidad & Tobago	12.14%	9.10%
46	Italy	12.07%	10.66%		Korea	15.12%	16.53%
47	Netherlands	7.89%	7.31%		Nepal	15.55%	8.59%
48	Norway	10.64%	8.00%		Pakistan	12.79%	7.88%
49	Switzerland	5.53%	4.54%		Thailand	10.86%	10.62%
50	Canada	8.99%	8.52%		Malawi	20.80%	17.62%
51	Japan	9.07%	6.16%		Zimbabwe	13.49%	12.01%
52	Ireland	9.43%	10.38%		Solomon Islands	15.89%	12.12%
53	Portugal	12.91%	10.79%		Western Samoa	12.90%	13.17%
54	Spain	10.38%	12.72%		Venezuela	28.47%	28.92%
55	Australia	9.15%	8.95%				

24. Mabelberry Fruit and Sawyer's Jam are two competing companies. An MBA student has done a calculation and found that the return on Sawyer's Jam stock is completely predictable once the return on Mabelberry Fruit stock is known:

$$r_{Sawyer's,t} = 40\% - 1.5 * r_{Mabelberry,t}$$

(a) Given the Mabelberry Fruit stock returns below, compute the Sawyer's Jam returns.

(b) Regress Mabelberry Fruit stock returns on those for Sawyer's Jam. Can you explain the R^2?

	A	B
2	**Year**	**Mabelberry Fruit stock return**
3	1990	30.73%
4	1991	15.00%
5	1992	-9.00%
6	1993	12.00%
7	1994	13.00%
8	1995	22.00%
9	1996	30.00%
10	1997	12.00%
11	1998	43.00%
12	1999	16.00%

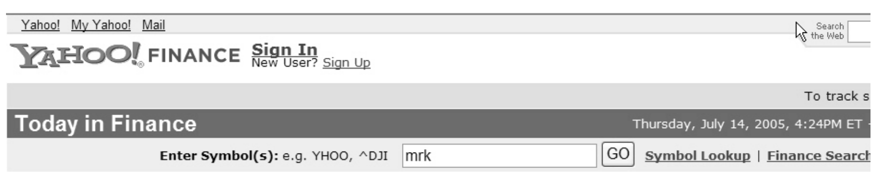
APPENDIX 12.1: DOWNLOADING DATA FROM YAHOO[12]

Yahoo provides free stock price data, which can be used to calculate returns. In this appendix we show you how to access this data and download it into Excel.

Step 1: Go to http://www.yahoo.com and click on Finance:

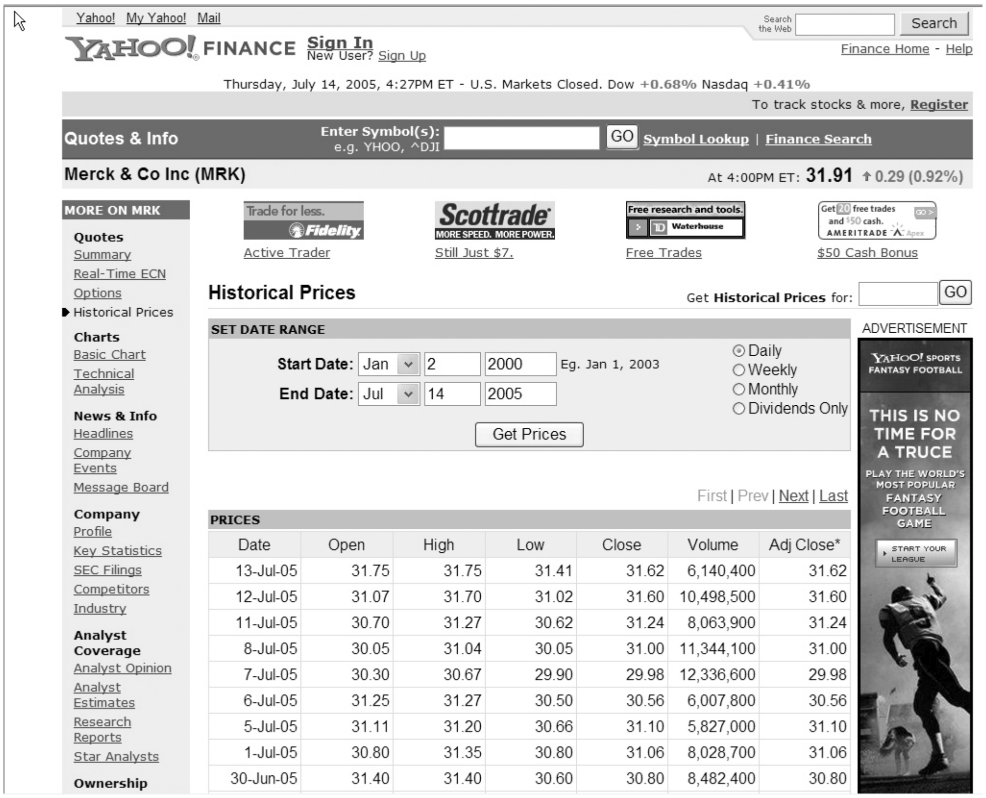
[12]Yahoo occasionally changes its interface; the information in this appendix is correct as of July 2006.

Step 2: In the "Enter symbol" box, put in the symbol for the stock you want to look up (we've put in MRK for Merck). You see that you can also look up symbols or put in multiple symbols. When you have put in the symbols, click on **Go**.



Step 3: This brings up the screen below. We choose **Historical Prices** to get Merck's price history.

Step 4: In the next screen, we indicated the time period and frequency for the data we want. Yahoo provides a table with stock prices, dividends, and an **Adjusted Closing Stock Price** that accounts for dividends and stock splits:

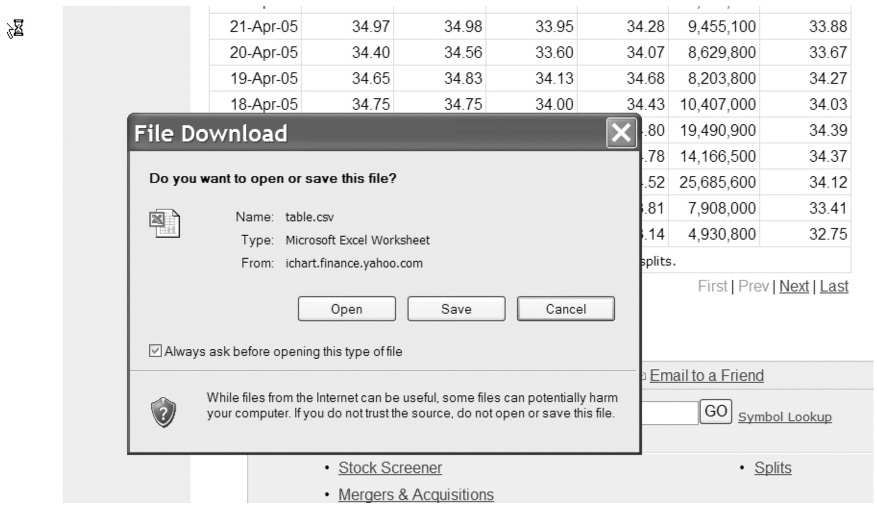


Date	Open	High	Low	Close	Volume	Adj Close*
13-Jul-05	31.75	31.75	31.41	31.62	6,140,400	31.62
12-Jul-05	31.07	31.70	31.02	31.60	10,498,500	31.60
11-Jul-05	30.70	31.27	30.62	31.24	8,063,900	31.24
8-Jul-05	30.05	31.04	30.05	31.00	11,344,100	31.00
7-Jul-05	30.30	30.67	29.90	29.98	12,336,600	29.98
6-Jul-05	31.25	31.27	30.50	30.56	6,007,800	30.56
5-Jul-05	31.11	31.20	30.66	31.10	5,827,000	31.10
1-Jul-05	30.80	31.35	30.80	31.06	8,028,700	31.06
30-Jun-05	31.40	31.40	30.60	30.80	8,482,400	30.80

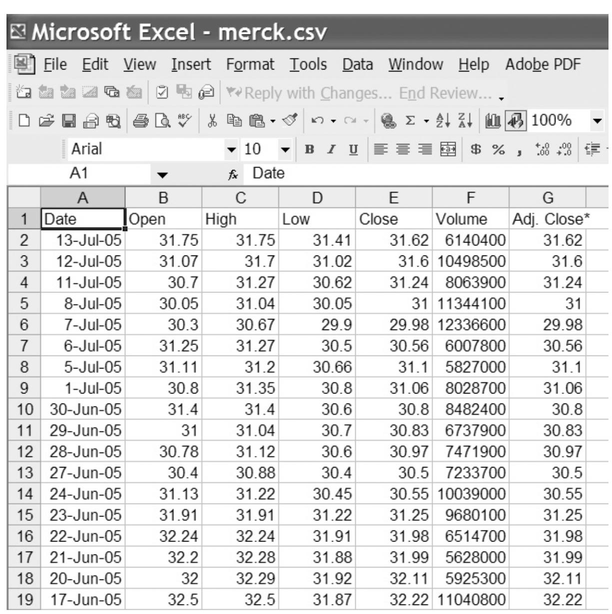
Step 5: The bottom of the above table allows you to download the data in spreadsheet format. In most browsers the Excel spreadsheet opens automatically (see results in Step 6):

21-Apr-05	34.97	34.98	33.95	34.28	9,455,100	33.88
20-Apr-05	34.40	34.56	33.60	34.07	8,629,800	33.67
19-Apr-05	34.65	34.83	34.13	34.68	8,203,800	34.27
18-Apr-05	34.75	34.75	34.00	34.43	10,407,000	34.03
15-Apr-05	35.24	36.26	34.76	34.80	19,490,900	34.39
14-Apr-05	34.60	34.97	34.53	34.78	14,166,500	34.37
13-Apr-05	33.81	35.30	33.52	34.52	25,685,600	34.12
12-Apr-05	33.05	33.82	32.95	33.81	7,908,000	33.41
11-Apr-05	33.45	33.59	33.05	33.14	4,930,800	32.75

Step 6: In the author's browser Yahoo offers to save a file called **Table.csv**. We changed the name of this file to **Merck.csv** and saved it on our hard disk.

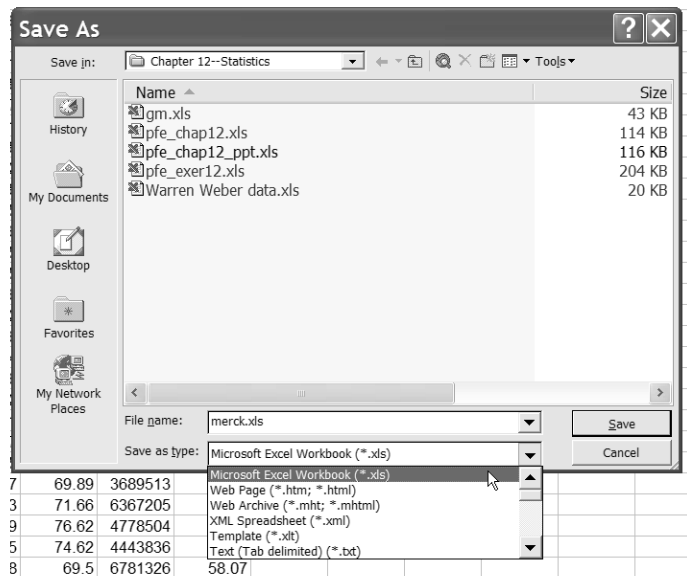


Step 7: The author's browser offered to open the file immediately (it will open as an Excel file). Here's the way opened Excel file looks. Note that only the adjusted stock prices are given.



	A	B	C	D	E	F	G
1	Date	Open	High	Low	Close	Volume	Adj. Close*
2	13-Jul-05	31.75	31.75	31.41	31.62	6140400	31.62
3	12-Jul-05	31.07	31.7	31.02	31.6	10498500	31.6
4	11-Jul-05	30.7	31.27	30.62	31.24	8063900	31.24
5	8-Jul-05	30.05	31.04	30.05	31	11344100	31
6	7-Jul-05	30.3	30.67	29.9	29.98	12336600	29.98
7	6-Jul-05	31.25	31.27	30.5	30.56	6007800	30.56
8	5-Jul-05	31.11	31.2	30.66	31.1	5827000	31.1
9	1-Jul-05	30.8	31.35	30.8	31.06	8028700	31.06
10	30-Jun-05	31.4	31.4	30.6	30.8	8482400	30.8
11	29-Jun-05	31	31.04	30.7	30.83	6737900	30.83
12	28-Jun-05	30.78	31.12	30.6	30.97	7471900	30.97
13	27-Jun-05	30.4	30.88	30.4	30.5	7233700	30.5
14	24-Jun-05	31.13	31.22	30.45	30.55	10039000	30.55
15	23-Jun-05	31.91	31.91	31.22	31.25	9680100	31.25
16	22-Jun-05	32.24	32.24	31.91	31.98	6514700	31.98
17	21-Jun-05	32.2	32.28	31.88	31.99	5628000	31.99
18	20-Jun-05	32	32.29	31.92	32.11	5925300	32.11
19	17-Jun-05	32.5	32.5	31.87	32.22	11040800	32.22

Step 8: It is advisable to use the Excel command **File|Save As** to save the file as a standard Excel file:



APPENDIX 12.2: WHY VARP INSTEAD OF VAR?

Throughout *Principles of Finance with Excel* we use the Excel functions **Varp** and **Stdevp** instead of their cousins **Var** and **Stdev**. This appendix briefly discusses this choice.

Recall that the definitions of these two functions relate to whether the data is taken from a *sample* or whether the data is the *whole population*. Suppose we have return data $\{r_1, r_2, \ldots, r_N\}$ for a stock. Then **Varp** is the population variance and **Var** is the sample variance:

$$\text{population variance} = Varp = \frac{1}{N}\sum_{t=1}^{N}(r_t - \bar{r})^2$$

$$\text{sample variance} = Var = \frac{1}{N-1}\sum_{t=1}^{N}(r_t - \bar{r})^2$$

There are two reasons why we choose **Varp** instead of **Var**: The first reason is that in most introductory statistics courses students are taught **Varp** (that is, divide by N) instead of **Var** (divide by $N - 1$). Thus, the choice made in *Principles of Finance with Excel* corresponds with what students have previously been taught.

The second reason for choosing **Varp** is that this choice makes the Excel **Slope** function consistent with the definition of β that we teach,

$$\beta_i = \frac{Covariance(r_{it}, r_{Mt})}{Variance(r_{Mt})}$$

To see this, reconsider the following example from Section 12.5 in which we calculate the β_{Mirage} for Mirage Resorts from monthly data for Mirage and for the S&P 500 Index:

	A	B	C	D
1	**WHY VARP INSTEAD OF VAR?**			
2	**Date**	**S&P 500 Index SPX**	**Mirage Resorts MIR**	
3	Jan-97	6.13%	16.18%	
4	Feb-97	0.59%	0.00%	
5	Mar-97	-4.26%	-15.42%	
6	Apr-97	5.84%	-5.29%	
7	May-97	5.86%	18.63%	
8	Jun-97	4.35%	5.76%	
9	Jul-97	7.81%	5.94%	
10	Aug-97	-5.75%	0.23%	
11	Sep-97	5.32%	12.35%	
12	Oct-97	-3.45%	-17.01%	
13	Nov-97	4.46%	-5.00%	
14	Dec-97	1.57%	-4.21%	
15	Jan-98	1.02%	1.37%	
16	Feb-98	7.04%	-0.54%	
17	Mar-98	4.99%	5.99%	
18	Apr-98	0.91%	-9.25%	
19	May-98	-1.88%	-5.67%	
20	Jun-98	3.94%	2.40%	
21	Jul-98	-1.16%	0.88%	
22	Aug-98	-14.58%	-30.81%	
23	Sep-98	6.24%	12.61%	
24	Oct-98	8.03%	1.12%	
25	Nov-98	5.91%	-12.18%	
26	Dec-98	5.64%	0.42%	
27				
28	Mirage β using **VarP**	1.4693	<-- =SLOPE(C3:C26,B3:B26)	
29		1.4693	<-- =COVAR(C3:C26,B3:B26)/VARP(B3:B26)	
30				
31	Mirage β using **Var**	1.4693	<-- =SLOPE(C3:C26,B3:B26)	
32		1.4080	<-- =COVAR(C3:C26,B3:B26)/VAR(B3:B26)	
33				
34	Market β using **Var**	1.0000	<-- =SLOPE(C3:C26,C3:C26)	
35		1.0000	<-- =COVAR(B3:B26,B3:B26)/VARP(B3:B26)	
36		0.9583	<-- =COVAR(B3:B26,B3:B26)/VAR(B3:B26)	
37				
38				
39				
40				
41				

The S&P index represents the market returns

The beta of the market should = 1. But using Covar(r_M,r_{Mirage})/Var(r_M) produces a beta < 1.

In cells B28 and B29 we compute the β_{Mirage} using the Excel function **Slope(C3:C26, B3:B26)** and the Excel functions **Covar(C3:C26,B3:B26)/VarP(B3:B26)**. These two definitions give the same (and the correct) answer. In cells B31 and B32 we compare the Excel **Slope** function to the answer given by **Covar(C3:C26,B3:B26)/Var(B3:B26)**. Note that the answers are different (the second answer is *incorrect*).

To drive home this point, we compute β_M in cells B34 to B36. The **Slope** function gives the correct answer, as does the definition **Covar(B3:B26,B3:B26)/VarP(B3:B26)**. However, using the function **Var** in cell B36 gives the *wrong* answer.

Conclusion: Best use **VarP** instead of **Var**![13]

[13]There's a slightly more cynical answer to the difference between **Var** and **Varp**. "If the difference between N and $N - 1$ ever matters to you, then you are probably up to no good anyway—e.g., trying to substantiate a questionable hypothesis with marginal data." This wonderful quote is from the book *Numerical Recipes* by William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vettering (Cambridge University Press, 1986, page 456).