# CHAPTER 9§

# Statistics

This chapter appears in the Survey text only.

THIS chapter attempts to distill the essential concepts of a full course in statistics into thirty-something pages. Thus, it is not an easy chapter, but it is also not as complex and painful as you might imagine.

## 9·1 Historical and Future Rates of Return

Finance is not philosophy—you have little else but history to guide us.

As an investor, your goal is to find the best possible investment portfolio. Easier said than done. What do you know about how a stock, say, IBM, will perform in the future? Not much. Your prime source of information about how IBM *will* perform is how it *did* perform. If it returned 10% per year over the last 10 years *on average*, maybe it is a good guess that it will return 10% over the next year, too. If it had a risk of plus or minus 20% per year, maybe it is a good guess that it will have a risk of plus or minus 20% over the next year, too. But, is historical performance really representative of future performance?

Don't take history too literally.

Clearly, it makes no sense to assume that future returns will be *exactly* the same as past returns. An investment in a particular six lotto numbers may have paid off big last year, but this does not mean that the exact same investment gamble will work again. More sensibly, you should look at the risk/reward characteristics of the *average* six lotto number investment, from which you would probably conclude that the average six number lotto investment is not a great bet.

For stocks, we use general historical characteristics to indicate future general characteristics.

Similarly, for stocks, it makes more sense to assume that future returns will be *only on average* like past returns in terms of risk and reward. It is not that we believe this to be exactly true, but it is usually our best guess *today*. Of course, we also know that the future will turn out different from the past—some firms will do better, others worse—but we generally have no better information than history. (If you can systematically estimate future risk and reward better than others, you are bound to become rich.)

Sometimes, even general historical characteristics (average) are obviously wrong.

Be warned, though. History is sometimes outright implausible as a predictor of future events. If you had played the lottery for 100 weeks and then won $1 million just by chance, a simple historical average rate of return would be $10,000 per week. Yes, it is a historical average, but it is not the right average *forward-looking*. (Of course, if wyouhad played the lottery 100 million weeks, we would almost surely come to the correct conclusion that playing the lottery is a gamble with a negative expected average rate of return.) Similarly, Microsoft or Wal-Mart are almost surely not going to repeat the spectacular historical stock return performances they experienced over the last 20 years. There are many examples when investors, believing history too much, made spectacularly wrong investment decisions. For example, in 1998-2000, Internet stocks had increased in value by more than 50% per year, and many investors believed that it was almost impossible to lose money on them. Of course, these investors, who believed historical Internet returns were indicative of future Internet returns, lost *all* their money over the following two years.

You have no better choice—so use it, but remain skeptical.

For the most part, the theory of investments—which is the subject of this part of our book—makes it easy on itself. It just *assumes* that you already know a stock's general risk/reward characteristics, and then proceeds to give you advice about what portfolio to choose, given that you already have the correct expectations. The problem of estimating means and variances remains *your* problem. Despite the problems with historical returns, I can only repeat: in many cases, stock's average historical risk-reward behavior is the best guidance you have. You can use this history, even though you should retain a certain skepticism—and perhaps even use common sense to adjust historical averages into more sensible forecasts for the future.

**Anecdote: Void Where Prohibited**

Persons pretending to forecast the future shall be considered disorderly under subdivision 3, section 901 of the criminal code and liable to a fine of $ 250 and/or six months in prison.

(Section 889, New York State Code of Criminal Procedure.)

## 9·2 The Data: Twelve Annual Rates of Returns

The goal of this chapter is to explain portfolios and stock returns under uncertainty. This is best done with a concrete example. Table 9.1 contains the actual twelve annual rates of returns from 1991 to 2002 for three possible investments: an S&P500 mutual fund, IBM stock, and Sony (ADR) shares.

The data example that is used throughout the Investments part of the book.

**Table 9.1:** Historical Annual Rates of Returns for S&P500, IBM, and Sony

| Year | $\tilde{r}_{S\&P500}$ | $\tilde{r}_{IBM}$ | $\tilde{r}_{Sony}$ | Year | $\tilde{r}_{S\&P500}$ | $\tilde{r}_{IBM}$ | $\tilde{r}_{Sony}$ |
|---|---|---|---|---|---|---|---|
| 1991 | +0.2631 | −0.2124 | −0.1027 | 1997 | +0.3101 | +0.3811 | +0.3905 |
| 1992 | +0.0446 | −0.4336 | −0.0037 | 1998 | +0.2700 | +0.7624 | −0.2028 |
| 1993 | +0.0706 | +0.1208 | +0.4785 | 1999 | +0.1953 | +0.1701 | +2.9681 |
| 1994 | −0.0154 | +0.3012 | +0.1348 | 2000 | −0.1014 | −0.2120 | −0.5109 |
| 1995 | +0.3411 | +0.2430 | +0.1046 | 2001 | −0.1304 | +0.4231 | −0.3484 |
| 1996 | +0.2026 | +0.6584 | +0.0772 | 2002 | −0.2337 | −0.3570 | −0.0808 |

Numbers are quoted in percent. You will be working with these returns throughout the rest of the investments part. (Sidenote: Sony is the SNE ADR.) Source: Yahoo!*Finance.*

You should first understand risk and reward, presuming it is now January 1, 2003. Although you are really interested in the returns of 2003 (and beyond), unfortunately all you have are these historical rates of return. You have to make the common assumption that historical returns are good indicators of future returns, because each year was an equally likely and therefore informative outcome drawn from an underlying statistical process. Formally, future returns are **random variables**, because their outcomes are not yet known. Recall from Sections 6·1 and 8·3.E that you can denote a random variable with a tilde over it, to distinguish it from an ordinary non-random variable, e.g.,

Pretend that historical rates of return are representative of future returns. Introduce tilde notation.

$$\tilde{r}_{S\&P500} \quad , \quad \tilde{r}_{IBM} \quad , \quad \tilde{r}_{Sony}$$

However, because you only know historical rates of return, you shall use the historical data series in place of the "tilde-d" future variables.

Conceptually, there is a big difference between average historical realizations and expected future realizations. Just because the long-run historical monthly rate of return average was, say, 10% does not mean that it will be 10% in the future. But, practically, there is often not much difference, because you have no choice but to pretend that the historical return series is representative of the distribution of future returns. To draw the distinction, statisticians often name the unknown expected value by a greek character, and the historical outcome that is used to estimate it by its corresponding English character. For example, $\mu$ would be the expected future mean, $m$ would be the historical mean; $\sigma$ would be the expected future standard deviation, $s$ would be the historical standard deviation. After having drawn this careful conceptual distinction, the statisticians then tell you that they will use the historical mean and historical standard deviation as stand-ins (estimators) for the future mean (the expected value) and standard deviation (the expected standard deviation). This is how it should be done, but it can become very cumbersome when dealing with many statistics for many variables. Because we shall mostly work with historical statistics and then immediately pretend that they are our expectations for the future, let us use a more casual notation: when we claim to compute $\mathcal{E}(\tilde{r})$, the notation of which would suggest an expected return, we really compute only the historical mean (unless otherwise stated). That is, in this finance book, we will keep the distinction rather vague.

Please realize that you are going to make a leap of faith here.

DIG DEEPER

> As with a history, you can think of the tilde as representing not just one month's outcome, but this distribution of historical outcomes. As of *today*, next month's rate of return can be anything. (We do not yet have just one number for it.)

## 9·3 Univariate Statistics

### 9·3.A. The Mean

Everyone knows how to compute an average.

If you had invested in a random year, what would you have expected to earn? The reward is measured by the single most important statistic, the expected rate of return (also called **mean** or **average** rate of return). You surely have computed this at one time or another, so let's just state that our means are

$$\mathcal{E}(\tilde{r}_{\text{S\&P500}}) = 0.101 \;\;,\;\; \mathcal{E}(\tilde{r}_{\text{IBM}}) = 0.154 \;\;,\;\; \mathcal{E}(\tilde{r}_{\text{Sony}}) = 0.242$$

Sony was clearly the best investment over these 12 years (mostly due to its spectacular performance in 1999), but IBM and the S&P500 did pretty well, too.

DIG DEEPER

> Chapter 5 already showed that the average rate of return is <u>not</u> the annualized rate of return. An investment in Sony beginning in 1999 for three years would have had a compound three-year rate of return of $(1+297\%)\cdot(1-51\%)\cdot(1-35\%) \approx 26.5\%$, which is 8.1% annualized. Its average annual rate of return is $[297\%+(-51\%)+(-35\%)]/3 \approx 70.3\%$.
>
> But, what causes the difference? It is the year-to-year volatility! if the rate of return were 8.1% each year without variation, the annualized and average rate of return would be the same. The year-to-year volatility negatively affects the annualized holding rates of return. For a given average rate of return, more volatility means less compound and thus less annualized rate of return.
>
> For purposes of forecasting a single year's return, assuming that each historical outcome was equally likely, you want to work with average rates of returns. For computing long-term holding period performance, you would want to work with compound rates of return.

### 9·3.B. The Variance and Standard Deviation

A Tale of three investments.

How can you measure portfolio risk? Intuitively, how does the risk of the following investment choices compare?

1. An investment in a bond that yields 10% per year for sure.
2. An investment that yields −15% half the time, and +35% half the time, but only once per year.
3. An investment in the S&P500.
4. An investment in IBM.
5. An investment in Sony.

You know that the first three investments have a mean rate of return of about +10% per annum. But the mean tells you nothing about the risk.

A naïve attempt at measuring spread fails.

You need a statistic that tells you how variable the outcomes are around the mean. Contestant #1 has no variability, so it is clearly safest. What about your other contestants? For contestant #2, half the time, the investment outcome is 25% below its mean (of +10%); the other half, it is 25% above its mean. Measuring outcomes relative to their means (as we have just done) is so common that it has its own name, **deviation from the mean**. Can you average the deviations from the mean to measure typical variability? Try it. With two years, and the assumption that each year is an equally likely outcome,

$$\text{Bad Variability Measure} = {}^1\!/_2 \cdot (-25\%) + {}^1\!/_2 \cdot (+25\%) = 0$$

The average deviation is zero, because the minus and plus cancel. Therefore, the simple average of the deviations is *not* a good measure of spread. You need a measure that tells you the typical variability is plus or minus 25%, not plus or minus 0%. Such a better measure must recognize that a negative deviation from the mean is the same as a positive deviation from the mean. The most common solution is to square each deviation from the mean in order to eliminate the "opposite sign problem." The average of these squared deviations is called the **variance**:

The root cause of the problem, and better alternatives: variance and standard deviation.

$$\mathcal{V}ar \;=\; {}^1\!/_2 \cdot (-25\%)^2 + {}^1\!/_2 \cdot (+25\%)^2 \;=\; {}^1\!/_2 \cdot 0.0625 + {}^1\!/_2 \cdot 0.0625 \;=\; 0.0625$$

You can think of the variance as the "average squared deviation from the mean." But the variance of 0.0625 looks nothing like the intuitive spread from the mean, which is plus or minus 25%. However, if you take the square root of the variance, you get the **standard deviation**,

The variance and standard deviation measure the expected spread of a random variable.

$$Sdv \;=\; \sqrt{\mathcal{V}ar} \;=\; \sqrt{0.0625} \;=\; 25\%$$

which has the intuitively pleasing correct order of magnitude of 25%. Although it is not really correct, it is often convenient to think of the standard deviation as the "average deviation from the mean." (It would be more correct to call it "the square root of the average squared deviation from the mean," but this is unwieldy.)

Aside from its uninterpretable value of 0.0625, there is a second and more important problem interpreting the meaning of a variance. (It did not show up in this example, because rates of return are unitless.) If you are interested in the variability of a variable that has units, like dollars or apples, the units of the variable are usually uninterpretable. For example, if you receive either \$10 or \$20, the deviation from the mean is either −\$5 or +\$5, and the variance is $(\$5)^2 = \$^2 25$, not \$25—the same way by which multiplying 2 meters by 2 meters becomes 4 square-meters, not 4 meters. Square-meters has a good interpretation (area); dollars-squared does not. The standard deviation takes the square root of the variance, and therefore returns to the same units (dollars) as the original series, $\sqrt{\$^2 25} = \$5$ in the example. Note also that I sometimes use "x%%" to denote $x \cdot (\%^2)$—otherwise, there may be confusion whether $x\%^2$ means $(x\%)^2$ or $x \cdot (\%^2)$. 1%% is $0.01 \cdot 0.01 = 0.0001$, and $\sqrt{1\%\%} = 1\%$.

SIDE NOTE

Because the standard deviation is just the square root of the variance, if the variance of one variable is higher than the variance of another variable, so is its standard deviation.

Standard deviation and variance have the same ordering.

**Table 9.2:** Deviations From the Mean for S&P500, IBM, and Sony

| Year | $\tilde{r}_{S\&P500}$ | $\tilde{r}_{IBM}$ | $\tilde{r}_{Sony}$ | Year | $\tilde{r}_{S\&P500}$ | $\tilde{r}_{IBM}$ | $\tilde{r}_{Sony}$ |
|---|---|---|---|---|---|---|---|
| 1991 | +0.1620 | −0.3661 | −0.3448 | 1997 | +0.2090 | +0.2273 | +0.1485 |
| 1992 | −0.0565 | −0.5874 | −0.2458 | 1998 | +0.1656 | +0.6086 | −0.4448 |
| 1993 | −0.0305 | −0.0330 | +0.2364 | 1999 | +0.0942 | +0.0163 | +2.7261 |
| 1994 | −0.1165 | +0.1474 | −0.1073 | 2000 | −0.2025 | −0.3658 | −0.7529 |
| 1995 | +0.2400 | +0.0892 | −0.1374 | 2001 | −0.2315 | +0.2693 | −0.5904 |
| 1996 | +0.1015 | +0.5046 | −0.1648 | 2002 | −0.3348 | −0.5105 | −0.3228 |
| | | | Mean (of Deviations) over all 12 years | | 0.0 | 0.0 | 0.0 |

For contestant #3, the S&P500, you must estimate the variability measures from the historical data series. Recall that to compute the variance, you subtract the mean from each outcome, square the deviations, and then average them. To compute the standard deviation, you then take the square-root of the variance. Table 9.2 does most of the hard work for you, giving you deviations from the mean for the S&P500, IBM, and Sony. Computing variances from these deviations is now straightforward: square and average. Alas, there is one nuisance complication: because there is a difference between historical realizations (which you have) and true expected

Sometimes-important nuisance: For historical data, do not divide the squared deviations by N, but by N-1.

future outcomes (which you do not have [we pretended to know this perfectly in the "-15%,+35%" example]), statisticians divide by $N-1$, not by $N$. Therefore, the estimated variance (divides by $N-1$) is a little bit higher than the average squared deviation from the mean (divides by $N$).

**SIDE NOTE**

The reason for this $N-1$ adjustment is that the future standard deviation is not actually known, but only estimated, given the historical realizations. This "extra uncertainty" is reflected by the smaller divisor, which inflates the uncertainty estimate.

The best intuition comes from a sample of only one historical data point: What would you believe the variability would be if you only know one realization, say 10%? In this case, you know nothing about variability. If you divided the average squared deviation (0) by $N$, the variance formula would indicate a zero variability. This is clearly wrong. If anything, you should be especially worried about variability, for you now know nothing about it. Dividing by $N-1 = 0$, i.e., $\mathcal{V}ar = 0/0$, indicates that estimating variability from one sample point makes no sense.

The division by $N$ rather than $N-1$ is not important when there are many historical sample data points, which is usually the case in finance. Thus, most of the time, you could use either method, although you should remain consistent. This book uses the $N-1$ statistical convention, if only because it allows checking computations against the built-in formulas in Excel, OpenOffice, or other statistical packages.

*Executing the formulas on historical data and dividing by N-1 yields variances and standard deviations.*

To obtain the variance of one investment series, square each deviation from the mean, add these squared terms, and dividing by 11 ($N-1$).

$$\mathcal{V}ar(\tilde{r}_{\text{S\&P500}}) = \frac{(+0.1620)^2 + (-0.0565)^2 + \cdots + (-0.3348)^2}{11} = 0.0362$$

$$\mathcal{V}ar(\tilde{r}_{\text{IBM}}) = \frac{(-0.3661)^2 + (-0.5874)^2 + \cdots + (-0.5105)^2}{11} = 0.1503$$

$$\mathcal{V}ar(\tilde{r}_{\text{Sony}}) = \frac{(-0.3448)^2 + (-0.2458)^2 + \cdots + (-0.3228)^2}{11} = 0.8149$$

$$\mathcal{V}ar(\tilde{r}) = \frac{[\tilde{r}_{t=0} - \mathcal{E}(\tilde{r})]^2 + [\tilde{r}_{t=1} - \mathcal{E}(\tilde{r})]^2 + \cdots + [\tilde{r}_{t=T} - \mathcal{E}(\tilde{r})]^2}{T-1}$$

The square roots of the variances are the standard deviations:

$$Sdv(\tilde{r}_{\text{S\&P500}}) = \sqrt{3.62\%} = 19.0\%$$

$$Sdv(\tilde{r}_{\text{IBM}}) = \sqrt{15.03\%} = 38.8\%$$

$$Sdv(\tilde{r}_{\text{Sony}}) = \sqrt{81.49\%} = 90.3\%$$

Return to your original question. If you were to line up your three potential investment choices—all of which offered about 10% rate of return—it appears that the S&P500 contestant #3 with its 19% risk is a safer investment than the "-15% or +35%" contestant #2 with its 25% risk. As for the other two investments, IBM with its higher 15%/year average rate of return is also riskier, having a standard deviation of "plus or minus" 38.8%/year. (Calling it "plus or minus" is a common way to express standard deviation.) Finally, Sony was not only the best performer (with its 24.2%/year mean rate of return), but it also was by far the riskiest investment. It had a whopping 90.3%/year standard deviation. (Like the mean, the large standard deviation is primarily caused by one outlier, the +297% rate of return in 1999.)

*Preview: Use of risk measures in finance.*

You will see that the mean and standard deviation play crucial roles in the study of investments—your ultimate goal will be to determine the portfolio that offers the highest expected reward for the lowest amount of risk. But mean and standard deviation make interesting statistics in themselves. From 1926 to 2002, a period with an inflation rate of about 3% per year, the annual risk and reward of some large asset-class investments were approximately

| Asset Class | $\mathcal{E}(\tilde{r})$ | $Sdv(\tilde{r})$ |
|---|---|---|
| Short-Term U.S. Government Treasury Bills | 4% | 3% |
| Long-Term U.S. Government Treasury Bonds | 5.5% | 10% |
| Long-Term Corporate Bonds | 6% | 9% |
| Large Firm Stocks | 10% | 20% |
| Small Firm Stocks | 15% | 30% |

Corporate bonds had more credit (i.e., default) risk than Treasury bonds, but were typically shorter-term than long-term government bonds, which explains their lower standard deviation. For the most part, it seems that higher risk and higher reward went hand-in-hand.

Solve Now!

**Q 9.1** Use a computer spreadsheet to confirm all numbers computed in this section.

**Q 9.2** The annual rates of return on the German DAX index were

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1991 | +0.1286 | 1994 | −0.0706 | 1997 | +0.4624 | 2000 | −0.0754 |
| 1992 | −0.0209 | 1995 | +0.0699 | 1998 | +0.1842 | 2001 | −0.1979 |
| 1993 | +0.4670 | 1996 | +0.2816 | 1999 | +0.3910 | 2002 | −0.4394 |

Compute the mean, variance, and standard deviation of the DAX.

## 9·4 Bivariate Statistics: Covariation Measures

### 9·4.A. Intuitive Covariation

Before we embark on more number-crunching, let us first find some intuitive examples of variables that tend to move together, variables that have nothing to do with one another, and variables that tend to move in opposite directions.

An example of a positive covariation.

Table 9.3 offers some such examples of covariation. For example, it is easier to score in basketball if you are 7 feet tall than if you are 5 feet tall. There is a positive covariation between individuals' heights and their ability to score. It is *not* perfect covariation—there are short individuals who can score a lot (witness John Stockton, the Utah Jazz basketball player, who despite a height of "only" 6-1 would almost surely score more points than the tallest students in my class), and there are many tall people who could not score if their lives depended on it. It is only *on average* that taller players score more. In this example, the reason for the positive covariation is direct causality—it is easier to hit the basket when you are as tall as the basket—but correlation need not come from causality. For example, there is also a positive covariation between shoe size and basketball scoring. It is not because bigger feet make it easier to score, but because taller people have both bigger shoe sizes and higher basketball scores. Never forget: causality induces covariation, but not vice-versa.

An example of a zero and a negative covariation.

Zero covariation usually means two variables have nothing to do with one another. For example, there is strong evidence that there is practically no covariation between gender and IQ. Knowing only the gender would not help you a bit in guessing the person's IQ, and vice-versa. (Chauvinists guessing wrong, however, tend to have lower IQs.) An example of negative covariation would be agility vs. weight. It is usually easier for lighter people to overcome the intrinsic inertia of mass, so they tend to be more agile: therefore, more weight tends to be associated with less agility.

**Table 9.3:** Covariation Examples

| Negative | Zero (or Low) | Positive |
|---|---|---|
| Agility vs. Weight | IQ vs. Gender | Height vs. Basketball Scoring |
| Wealth vs. Disease | Wealth vs. Tail when flipping coin | Wealth vs. Longevity |
| Age vs. Flexibility | Age vs. Blood Type | Age vs. Being CEO |
| Snow vs. Temperature | Sunspots vs. Temperature | Grasshoppers vs. Temperature |
| Your Net Returns vs. Broker Commissions Paid | IBM Returns in 1999 vs. Exxon Returns in 1986 | Returns on S&P500 vs. Returns on IBM |

Personal statistics (such as weight) apply only to adults. Returns are rates of return on stock investments, net of commissions.

## 9·4.B. Covariation: Covariance, Correlation, and Beta

A measure of dependence should be positive when two variables tend to move together, negative if they tend to move in opposite directions.

Your goal now is to find measures of covariation that are positive when two variables tend to move together; that are zero when two variables have nothing to do with one another; and that are negative when one variable tends to be lower when the other variable tends to be higher. We will consider three possible measures of covariation: covariance, correlation, and beta. Each has its advantages and disadvantages.

### Illustrative Data Series

Work with nine made-up data series.

Start with the nine data series in Table 9.4, the returns on nine assets that I have made up. Let us use asset A as our base asset and consider how assets C through J relate to A. You want to ask such questions as “if data series A were the rate of return on the S&P500, and data series C were the rate of return on IBM, then how do the two return series covary?” This question is important, because it will help you determine investment opportunities that have lower risk.

Covariation is best to understand with an extreme, albeit absurd example. But you need two measures of covariation, not just one.

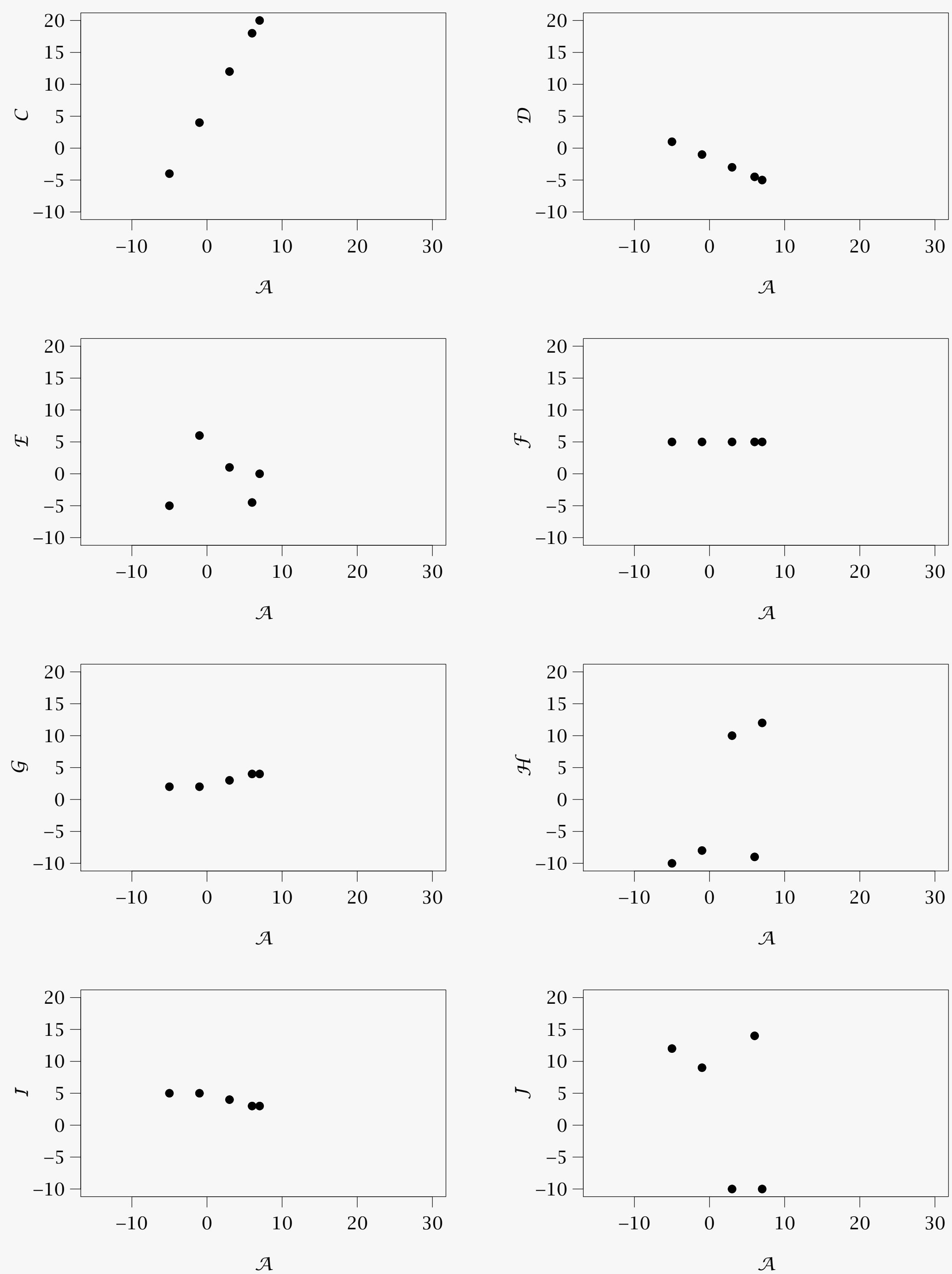
How does each series covaries with A? Doing this graphically makes it easier, so Figure 9.1 plots the points. If you look at it, you can see that you shall need more than just one covariation statistic: you need one statistic that tells you how much two variables are related (e.g., whether A has more effect on G or H); and one statistic that tells you how reliable this relationship is (e.g., whether knowing A gives you much confidence to predict G or H). Your first task is to draw lines into each of the eight graphs in Figure 9.1 that best fits the points (do it!)—and then to stare at the points and your lines. How steep are your lines (the relationships) and how reliably do the points cluster around your line? Before reading on, your second task is to write into the eight graphs in Figure 9.1 what your intuition tells you the relation and the reliability are.

Here is what I see.

Figure 9.1 seems to show a range of different patterns:

- C and D are strongly related to A. Actually, C is just $2 \cdot (A + 3)$; D is just $-C/2$, so these two relationships are also perfectly reliable.
- E and F have no relationship to A. Though random, there is no pattern connecting A and E. F is 5 regardless of what A is.
- G and H both tend to be higher when A is higher, but in different ways. There is more reliability in the relation between A and G, but even a large increase in A predicts a G that

**Figure 9.1:** C Through J as functions of A




The five observations are marked by circles. *Can you draw a well fitting line? Which series have relations with A? What sign? Which series have reliable relations with A?*

**Table 9.4:** Illustrative Rates of Return Time Series on Nine Assets

| | | | Portfolio (or Asset or Security) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observation | A | C | D | E | F | G | H | I | J |
| Year 1 | −5 | −4 | +1.0 | −10 | +5 | 2 | −10 | +5 | +12 |
| Year 2 | +6 | +18 | −4.5 | −9 | +5 | 4 | −9 | +3 | +14 |
| Year 3 | +3 | +12 | −3.0 | +2 | +5 | 3 | +10 | +4 | −10 |
| Year 4 | −1 | +4 | −1.0 | +12 | +5 | 2 | −8 | +5 | +9 |
| Year 5 | +7 | +20 | −5.0 | 0 | +5 | 4 | +12 | +3 | −10 |
| Mean | +2 | +10 | −2.5 | −1 | +5 | +3 | −1 | +4 | +3 |
| *Var* | 25 | 100 | 6.25 | 81 | 0 | 1 | 121 | 1 | 144 |
| *Sdv* | 5 | 10 | 2.5 | 9 | 0 | 1 | 11 | 1 | 12 |

*Which rate of return series (among portfolios C through J) had low and which had high covariation with the rate of return series of Portfolio A?*

is only a little higher. In contrast, there is less reliability in how A and H covary, but even a small increase in A predicts a much higher H.

➤ Figures I and J repeat the G/H pattern, but for negative relations.

SIDE NOTE

I cheated in not using my eyeballs to draw lines, but in using the technique of "ordinary least squares" line fitting in Figure 9.3, instead. The lines make it even clearer that when A is high, C, G, and H tend to be high, too; but when A is high, D, I, and J tend to be low. And neither E nor F seem to covary with A. (You will get to compute the slope of this line—the "beta"—later.)

How covariance really works: quadrants above/below means.

Of course, visual relationships are in the eye of the beholder. You need something more objective that both of us can agree on. Here is how to compute a precise measure. The first step is to determine the means of each $X$ and each $Y$ variable and mark these into the figure—which is done for you in Figure 9.2. The two means divide your data into four quadrants. Now, intuitively, points in the northeast or southwest quadrants (in white) suggest a positive covariation; points that are in the northwest or southeast quadrants (in red) suggest a negative covariation. In other words, the idea of all of the covariation measures is that two series, call them $X$ and $Y$, are positively correlated

➤ when $X$ tends to be above its mean, $Y$ also tends to be above its mean (upper right quadrant); and

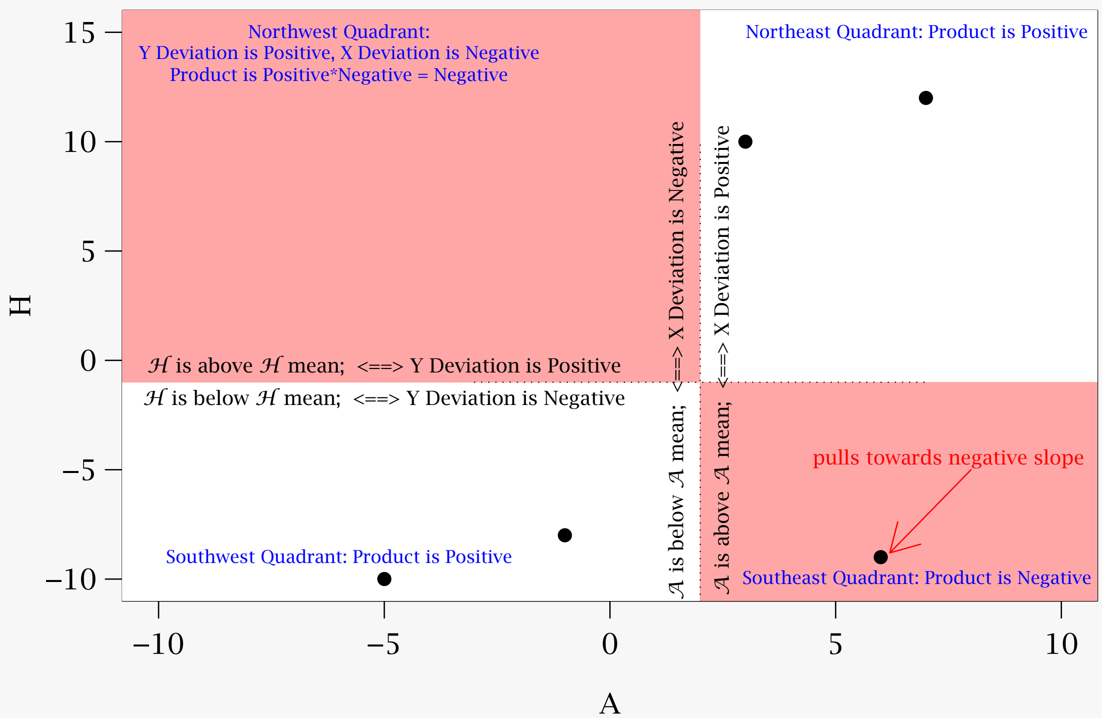
➤ when $X$ tends to be below its mean, $Y$ also tends to be below its mean (lower left quadrant).

And $X$ and $Y$ are negatively correlated

➤ when $X$ tends to be above its mean, $Y$ tends to be below its mean (lower right quadrant); and

➤ when $X$ tends to be below its mean, $Y$ tends to be above its mean (upper left quadrant).

**Figure 9.2:** H as Function of A




Points in the red quadrants pull the overall covariance statistics into a negative direction. Points in the white quadrants pull the overall covariance statistics into a positive direction.

## Covariance

How can you make a positive number for every point that is either above both the $X$ and $Y$ means, or both below the $X$ and $Y$ means, and a negative number for every point that is above one mean and below the other? Easy! First, you measure each data point in terms of its distance from its mean, so you subtract the mean from each data point, as in Table 9.5. Points in the northeast quadrant are above both means, so both net-of-mean values are positive. Points in the southwest quadrant are below both means, so both net-of-mean values are negative. Points in the other two quadrants have one positive and one negative net-of-mean number. Now, you know that the product of either two positive or two negative numbers is positive, and the product of one positive and one negative number is negative. If you multiply your deviations from the mean, the product has a positive sign if it is in the upper-right and lower-left quadrants (the deviations from the mean are either both positive or both negative), and a negative sign if it is in the upper-left and lower-right quadrants (only one deviation from the mean is positive, the other is negative). A point that has a positive product pulls towards positive covariation, whereas a negative product pulls towards negative covariation.

The main idea is to think about where points lie, within the quadrants in the figure. Multiply the deviations from the mean to get the right sign for each data point relative to each quadrant.

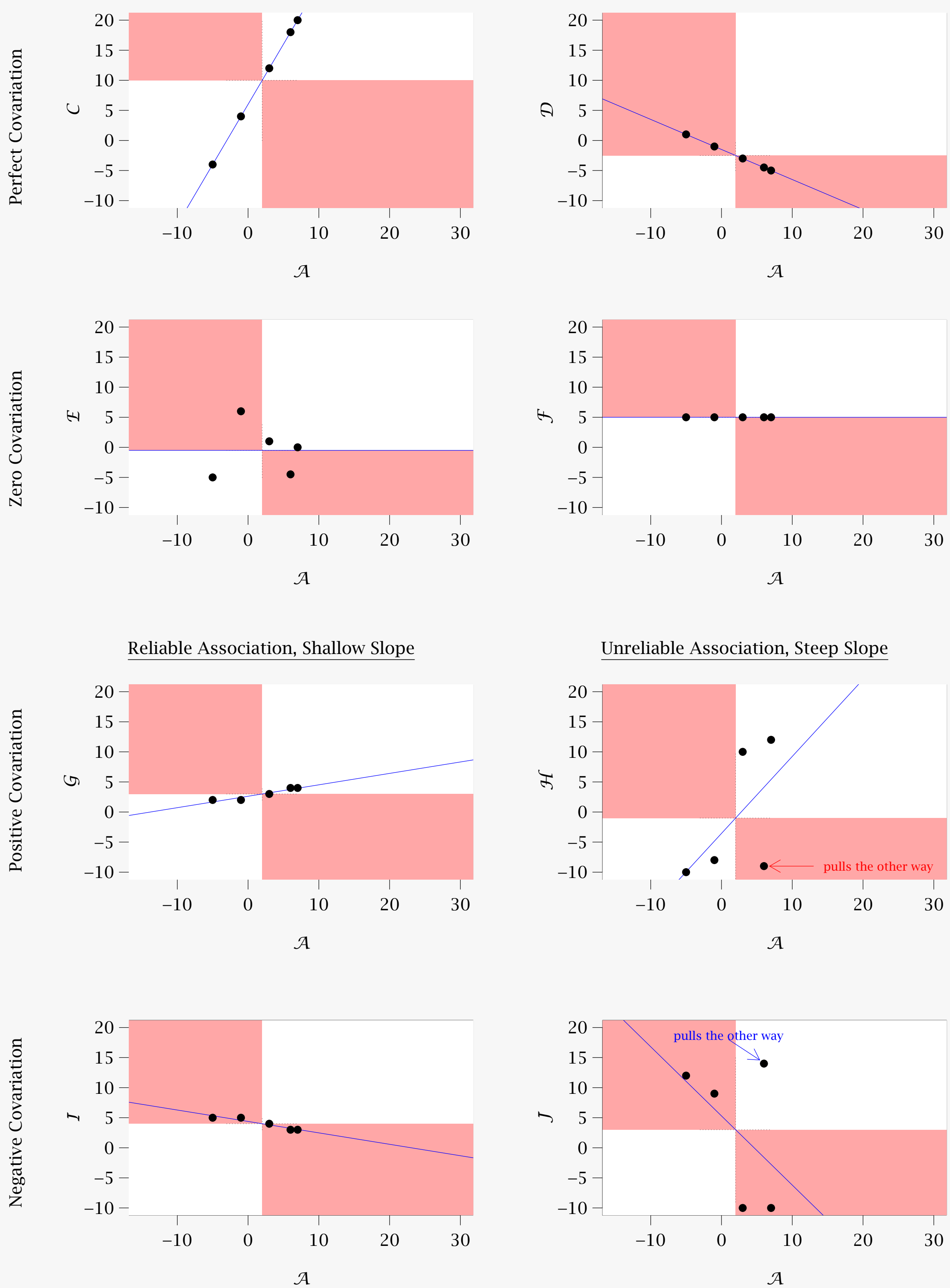
You want to know the average pull. This is the **covariance**—the average of these products of two variables' deviations from their own means. Try it for A and C (again, following the standard method of dividing by $N-1$ rather than $N$):

The covariance is the average of the multiplied deviations from their means.

$$Cov(\text{A},\text{C}) = \frac{(-7)\cdot(-14) + (+4)\cdot(+8) + (+1)\cdot(+2) + (-3)\cdot(-6) + (+5)\cdot(+10)}{4} = 50$$

$$Cov(\text{A},\text{C}) = \frac{(a_1-\bar{a})\cdot(b_1-\bar{b}) + \cdots + (a_5-\bar{a})\cdot(b_5-\bar{b})}{N-1}$$

**Figure 9.3:** B Through I as functions of A, Lines Added




The five observations are marked by circles. The areas north, south, east, and west of the $X$ and $Y$ means are now marked. A cross with arm lengths equal to one standard deviation is also placed on each figure.

*Which points push the relationship to be positive, which points push the relationship to be negative?*

**Table 9.5:** Illustrative Rates of Return Time Series on Nine Assets, De-Meaned

| | Rates of Returns on Portfolios (or Assets or Securities) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Observation | A | C | D | E | F | G | H | I | J |
| Year 1 | −7 | −14 | +3.5 | −9 | 0 | −1 | −9 | +1 | +9 |
| Year 2 | +4 | +8 | −2.0 | −8 | 0 | +1 | −8 | −1 | +11 |
| Year 3 | +1 | +2 | −0.5 | +3 | 0 | 0 | +11 | 0 | −13 |
| Year 4 | −3 | −6 | +1.5 | +13 | 0 | −1 | −7 | +1 | +6 |
| Year 5 | +5 | +10 | −2.5 | +1 | 0 | +1 | +13 | −1 | −13 |
| $\mathcal{E}(\tilde{r})$ | 0 | 0 | 0.0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $\mathcal{V}ar(\tilde{r})$ | 25 | 100 | 6.25 | 81 | 0 | 1 | 121 | 1 | 144 |
| $Sdv(\tilde{r})$ | 5 | 10 | 2.5 | 9 | 0 | 1 | 11 | 1 | 12 |

It will be easier to work with the series from Table 9.4 if you first subtract the mean from each series.

Note how in this A vs. C relationship, each term in the sum is positive, and therefore pulls the average (the covariance) towards being a positive number. You can see this in the figure, because each and every point lies in the two "positivist" quadrants. Repeat this for the remaining series:

$$Cov(A,D) = \frac{(-7)\cdot(+3.5) + (+4)\cdot(-2) + (+1)\cdot(-0.5) + (-3)\cdot(+1.5) + (+5)\cdot(-2.5)}{4} = -12.5$$

$$Cov(A,E) = \frac{(-7)\cdot(-9) + (+4)\cdot(-8) + (+1)\cdot(+3) + (-3)\cdot(+13) + (+5)\cdot(+1)}{4} = 0$$

$$Cov(A,F) = \frac{(-7)\cdot(0) + (+4)\cdot(0) + (+1)\cdot(0) + (-3)\cdot(0) + (+5)\cdot(0)}{4} = 0$$

$$Cov(A,G) = \frac{(-7)\cdot(-1) + (+4)\cdot(+1) + (+1)\cdot(0) + (-3)\cdot(-1) + (+5)\cdot(+1)}{4} = 4.75$$

$$Cov(A,H) = \frac{(-7)\cdot(-9) + (+4)\cdot(-8) + (+1)\cdot(+11) + (-3)\cdot(-7) + (+5)\cdot(+13)}{4} = 32$$

$$Cov(A,I) = \frac{(-7)\cdot(+1) + (+4)\cdot(-1) + (+1)\cdot(0) + (-3)\cdot(+1) + (+5)\cdot(-1)}{4} = -4.75$$

$$Cov(A,J) = \frac{(-7)\cdot(+9) + (+4)\cdot(+11) + (+1)\cdot(-13) + (-3)\cdot(+6) + (+5)\cdot(-13)}{4} = -28.75$$

$$Cov(X,Y) = \frac{(x_1-\bar{x})\cdot(y_1-\bar{y}) + \cdots + (x_5-\bar{x})\cdot(y_5-\bar{y})}{N-1}$$

Having computed the covariances, look at Figure 9.3. A and D, A and I, and A and J covary negatively on average; A and C, A and G, and A and H covary positively; and A and E, and A and F have zero covariation. Now, take a look at the A vs. H figure (also in Figure 9.2) again: there is one lonely point in the lower-right quadrant, marked with an arrow. It tries to pull the line into a negative direction. In the A vs. H covariance computation, this point is the $(+4)\cdot(-1)$ term, which is the only negative component in the overall sum. If this one point had not been in the data, the association between A and H would have been more strongly positive than 4.75.

Covariance gives the right sign, but not much more. It is often abbreviated as a sigma with two subscripts.

The covariance tells you the sign (whether a relationship is positive or negative), but its magnitude is difficult to interpret—just as you could not really interpret the magnitude of the variance. Indeed, the covariance not only shares the weird squared-units problem with the variance, but the covariance of a variable with itself *is* the variance! This can be seen from the definitions: Both multiply each historical outcome deviation by itself, and then divide by the same number,

$$Cov(X,X) = \frac{(x_1-\bar{x})\cdot(x_1-\bar{x}) + \cdots + (x_N-\bar{x})\cdot(x_N-\bar{x})}{N-1}$$

$$= \frac{(x_1-\bar{x})^2 + \cdots + (x_N-\bar{x})^2}{N-1} = \mathcal{V}ar(X)$$

And, just like the variance is needed to compute the standard deviation, the covariance is needed to compute the next two covariation measures (correlation and beta). The covariance statistic is so important and used so often that the Greek letter sigma ($\sigma$) with two subscripts has become the standard abbreviation:

$$\text{Covariance between } X \text{ and } Y\text{:} \quad Cov(X,Y) \qquad \sigma_{X,Y}$$

$$\text{Variance of } X\text{:} \quad Var(X) \qquad \sigma_{X,X} = \sigma_X^2$$

These are sigmas with two subscripts. If you use only one subscript, you mean the standard deviation:

$$\text{Standard Deviation of } X\text{:} \quad Sdv(X) \qquad \sigma_X$$

This is easy to remember if you think of two subscripts as the equivalent of multiplication (squaring), and of one subscript as the equivalent of square-rooting.

## Correlation

Correlation is closely related to, but easier to interpret than Covariance.

To better interpret the covariance, you need to somehow normalize it. A first normalization of the covariance gives the **correlation**. It divides the covariance by the standard deviations of both variables. Applying this formula, you can compute

$$\begin{aligned}
Correlation(A,C) &= \frac{Cov(A,C)}{Sdv(A)\cdot Sdv(C)} = \frac{+50}{10\cdot 5} = +1.00\\
Correlation(A,D) &= \frac{Cov(A,D)}{Sdv(A)\cdot Sdv(D)} = \frac{-12.5}{5\cdot 2.5} = -1.00\\
Correlation(A,E) &= \frac{Cov(A,E)}{Sdv(A)\cdot Sdv(E)} = \frac{0}{5\cdot 9} = \pm 0.00\\
Correlation(A,F) &= \frac{Cov(A,F)}{Sdv(A)\cdot Sdv(F)} = \frac{0}{5\cdot 0} = \text{not defined}\\
Correlation(A,G) &= \frac{Cov(A,G)}{Sdv(A)\cdot Sdv(G)} = \frac{4.75}{5\cdot 1} = +0.95\\
Correlation(A,H) &= \frac{Cov(A,H)}{Sdv(A)\cdot Sdv(H)} = \frac{32}{5\cdot 11} = +0.58\\
Correlation(A,I) &= \frac{Cov(A,I)}{Sdv(A)\cdot Sdv(I)} = \frac{-4.75}{5\cdot 1} = -0.95\\
Correlation(A,J) &= \frac{Cov(A,J)}{Sdv(A)\cdot Sdv(J)} = \frac{-28.75}{5\cdot 12} = -0.48\\
Correlation(X,Y) &= \frac{Cov(X,Y)}{Sdv(X)\cdot Sdv(Y)}
\end{aligned} \tag{9.1}$$

The correlation measures the reliability of the relationship between two variables. A higher *absolute* correlation means more reliability, regardless of the strength of the relationship (slope). The nice thing about the correlation is that it is always between −100% and +100%. Two variables that have a correlation of 100% always perfectly move in the same direction, two variables that have a correlation of –100% always perfectly move in the opposite direction, and two variables that are independent have a correlation of 0%. This makes the correlation very easy to interpret. The correlation is unit-less, regardless of the units of the original variables themselves, and is often abbreviated with the Greek letter rho ($\rho$). The perfect correlations between A and C or D tell you that all points lie on straight lines. (Verify this visually in Figure 9.3!) The correlations between A and G (95%) and the correlations between A and I (−95%) are almost as strong: the points almost lie on a line. The correlation between A and H, and the correlation between A and J are weaker: the points do not seem to lie on a straight line, and knowing A does not permit you to *perfectly* predict H or J.

If two variable are always acting identically, they have a correlation of 1. Therefore, you can determine the maximum covariance between two variables:

$$1 = \frac{Cov(\tilde{X},\tilde{Y})}{Sdv(\tilde{X}) \cdot Sdv(\tilde{Y})} \iff Cov(\tilde{X},\tilde{Y}) = Sdv(\tilde{X}) \cdot Sdv(\tilde{Y})$$

It is mathematically impossible for the absolute value of the covariance to exceed the product of the two standard deviations.

SIDE NOTE

## Beta

Beta is yet another covariation measure, and has an interesting graphical interpretation.

The correlation cannot tell you that A has more pronounced influence on C than on D: although both correlations are perfect, if A is higher by 1, your prediction of C is higher by 2; but if A is higher by 1, your prediction of D is lower by only −0.5. You need a measure for the slope of the best-fitting line that you would draw through the points. Your second normalization of the covariance does this: it gives you this slope, the **beta**. It divides the covariance by the variance of the *X* variable (here, A), i.e., instead of one dividing by the standard deviation of *Y* (as in the correlation), it divides a second time by a standard deviation of *X*:

$$\beta_{C,A} = \frac{Cov(A,C)}{Sdv(A) \cdot Sdv(A)} = \frac{+50}{5 \cdot 5} = +2.00$$

$$\beta_{D,A} = \frac{Cov(A,D)}{Sdv(A) \cdot Sdv(A)} = \frac{-12.5}{5 \cdot 5} = -0.50$$

$$\beta_{E,A} = \frac{Cov(A,E)}{Sdv(A) \cdot Sdv(A)} = \frac{0}{5 \cdot 5} = \pm 0.00$$

$$\beta_{F,A} = \frac{Cov(A,F)}{Sdv(A) \cdot Sdv(A)} = \frac{0}{5 \cdot 5} = \pm 0.00$$

$$\beta_{G,A} = \frac{Cov(A,G)}{Sdv(A) \cdot Sdv(A)} = \frac{4.75}{5 \cdot 5} = +0.19$$

$$\beta_{H,A} = \frac{Cov(A,H)}{Sdv(A) \cdot Sdv(A)} = \frac{32}{5 \cdot 5} = +1.28$$

$$\beta_{I,A} = \frac{Cov(A,I)}{Sdv(A) \cdot Sdv(A)} = \frac{-4.75}{5 \cdot 5} = -0.19$$

$$\beta_{J,A} = \frac{Cov(A,J)}{Sdv(A) \cdot Sdv(A)} = \frac{-28.75}{5 \cdot 5} = -1.15$$

$$\beta_{Y,X} = \frac{Cov(X,Y)}{Sdv(X) \cdot Sdv(X)} = \frac{Cov(X,Y)}{\mathcal{V}ar(X)}$$

The first subscript on beta denotes the variable on the *Y* axis, while the second subscript on beta denotes the variable on the *X* axis. It is the latter that provides the variance (the divisor). Beta got its name from the fact that the most common way to write down the formula for a line is $y = \alpha + \beta \cdot x$, and the best-fitting line slope is exactly what beta is. Unlike correlations, betas are not limited to any range. Beta values of +1 and −1 denote the diagonal lines, beta values of 0 and ∞ denote the horizontal and vertical line. Inspection of Figure 9.3 shows that the slope of A vs. C is *2 to 1*, while the slope of A vs. D is shallower *1 to -2*. This is exactly what beta tells us: $\beta_{C,A}$ is 2.0, while $\beta_{D,A}$ is −0.5. Unlike the correlation, beta cannot tell you whether your line fits perfectly or imperfectly. But, unlike the correlation, beta can tell you how much you should change your prediction of *Y* if the *X* values change. And unlike correlation and covariance, the order of the two variables matters in computing beta. For example, $\beta_{G,A} = 0.19$ is not $\beta_{A,G}$:

$$\beta_{A,\,G} = \frac{Cov(A,G)}{Sdv(G) \cdot Sdv(G)} = \frac{5.00}{1 \cdot 1} = 5$$

In a statistical package, beta can be obtained either by computing covariances and variances and then dividing the two; or by running a **Linear Regression**, in which the dependent variable is the *Y* variable and the independent variable is the *X* variable.

$$Y = \alpha + \beta \cdot X + \epsilon$$

Both methods yield the same answer.

DIG DEEPER

When it comes to stock returns, you really want to know the future slope, although you usually only have the historical slope.

And, as with all other statistical measures, please keep in mind that you are usually computing a historical beta (line slope), although you usually are really interested in the future beta (line slope)!

### Summary of Covariation Measures

Table 9.6 summarizes the three covariation measures.

**Table 9.6:** Comparison of Covariation Measures

| | Units | Magnitude | Order of Variables | computed as | Measures |
|---|---|---|---|---|---|
| Covariance | squared | practically meaningless | irrelevant | $\sigma_{X,Y}$ | No Intuition |
| Correlation | unit-less | between −1 and +1 | irrelevant | $\sigma_{X,Y} / (\sigma_X \cdot \sigma_Y)$ | Reliability |
| beta (Y,X) | unit-less | meaningful (slope) | important | $\sigma_{X,Y} / \sigma_{X,X}$ | Slope |
| beta (X,Y) | unit-less | meaningful (slope) | important | $\sigma_{X,Y} / \sigma_{Y,Y}$ | Slope |

All covariation measures share the same sign. If one is positive (negative), so are all others. Recall that $\sigma_{X,X} = \sigma_X^2$, which must be positive.

Figure 9.4 summarizes everything that you have learned about the covariation of your series. It plots the data points, the quadrants, the best fitting lines, and a text description of the three measures of covariation.

## 9·4.C. Computing Covariation Statistics For The Annual Returns Data

Applying the covariance formula to the historical data.

Now back to work! It is time to compute the covariance, correlation, and beta for your three investment choices, S&P500, IBM, and Sony. Return to the deviations from the means in Table 9.2. As you know, to compute the covariances, you add the products of the net-of-mean observations and divide by $(T - 1)$—tedious, but not difficult work:

$$Cov(\tilde{r}_{\text{S\&P500}}, \tilde{r}_{\text{IBM}}) = \frac{(+0.162) \cdot (-0.366) + \cdots + (-0.335) \cdot (-0.511)}{11} = 0.0330$$

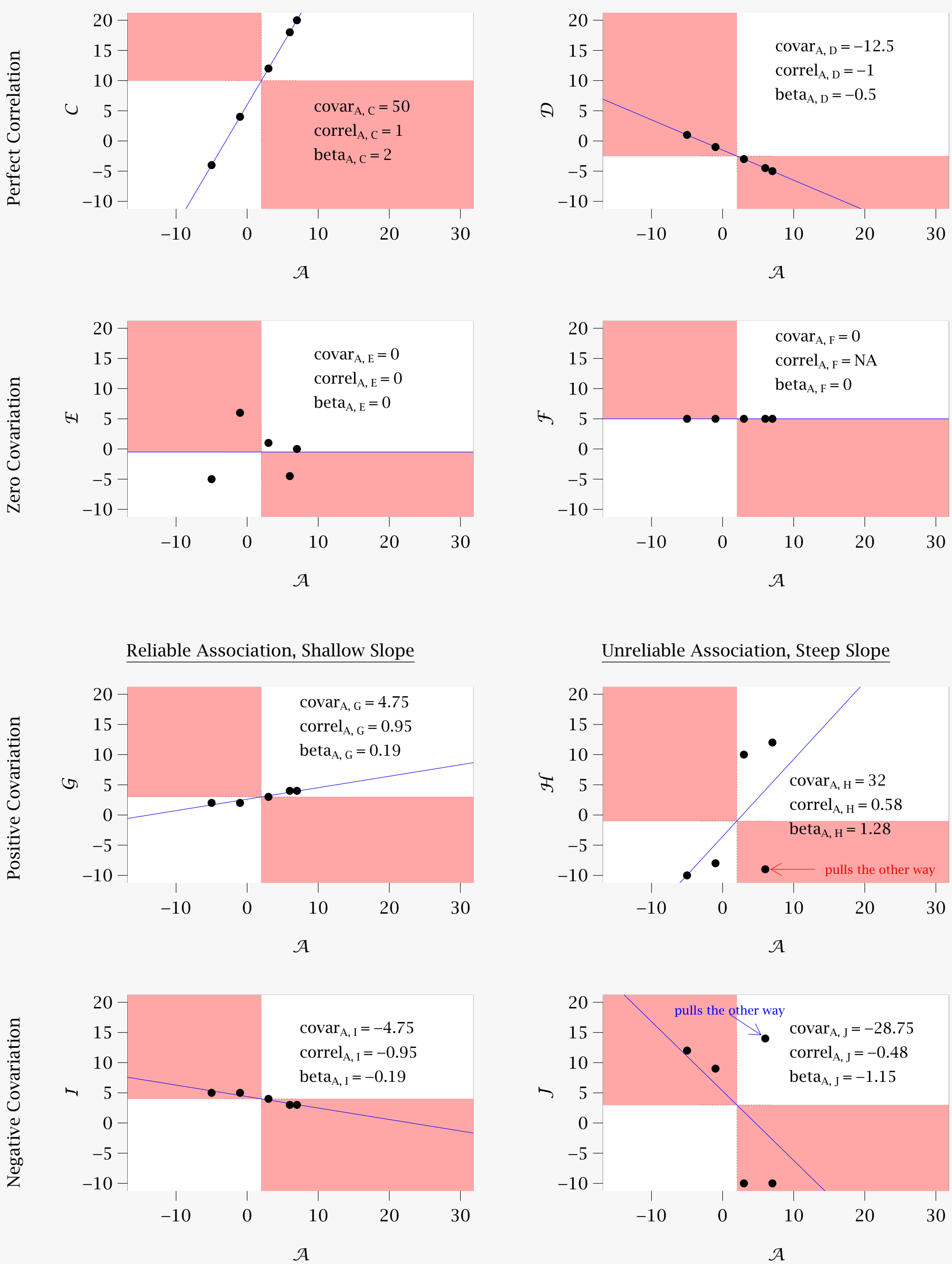
$$Cov(\tilde{r}_{\text{S\&P500}}, \tilde{r}_{\text{Sony}}) = \frac{(+0.162) \cdot (-0.345) + \cdots + (-0.335) \cdot (-0.323)}{11} = 0.0477$$

$$Cov(\tilde{r}_{\text{IBM}}, \tilde{r}_{\text{Sony}}) = \frac{(-0.366) \cdot (-0.345) + \cdots + (-0.511) \cdot (-0.323)}{11} = 0.0218$$

$$Cov(\tilde{r}_{\text{i}}, \tilde{r}_{j}) = \frac{(\tilde{r}_{\text{i},s=1} - \mathcal{E}(\tilde{r}_{\text{i}})) \cdot (\tilde{r}_{\text{j},s=1} - \mathcal{E}(\tilde{r}_{\text{j}})) + \cdots + (\tilde{r}_{\text{i},s=T} - \mathcal{E}(\tilde{r}_{\text{i}})) \cdot (\tilde{r}_{\text{j},s=T} - \mathcal{E}(\tilde{r}_{\text{j}}))}{T - 1} \tag{9.2}$$

All three covariance measures are positive. You know from the discussion on Page 199 that, aside from their signs, the covariances are almost impossible to interpret. Therefore, now

**Figure 9.4:** C Through J as functions of A with lines and text




The five observations are marked by circles. The areas north, south, east, and west of the $X$ and $Y$ means are now marked. A cross with arm lengths equal to one standard deviation is also placed on each figure.

compute the correlations, your measure of how well the best-fitting line fits the data. The correlations are the covariances divided by the two standard deviations:

$$\begin{aligned}
Correlation(\tilde{r}_{\text{S\&P500}}, \tilde{r}_{\text{IBM}}) &= \frac{3.30\%}{19.0\% \cdot 38.8\%} = 44.7\% \\
Correlation(\tilde{r}_{\text{S\&P500}}, \tilde{r}_{\text{Sony}}) &= \frac{4.77\%}{19.0\% \cdot 90.3\%} = 27.8\% \\
Correlation(\tilde{r}_{\text{IBM}}, \tilde{r}_{\text{Sony}}) &= \frac{2.18\%}{38.8\% \cdot 90.3\%} = 6.2\% \\
Correlation(\tilde{r}_{\text{i}}, \tilde{r}_{\text{j}}) &= \frac{Cov(\tilde{r}_{\text{i}}, \tilde{r}_{\text{j}})}{Sdv(\tilde{r}_{\text{i}}) \cdot Sdv(\tilde{r}_{\text{j}})}
\end{aligned}$$

The S&P500 has correlated much more with IBM, than the S&P500 has correlated with Sony (or IBM with Sony). This makes intuitive sense. Both S&P500 and IBM are U.S. investments, while Sony is a stock that is trading in an entirely different market.

Applying the beta formula to the historical data.

Finally, you might want to compute the beta of $\tilde{r}_{\text{Sony}}$ with respect to the $\tilde{r}_{\text{S\&P500}}$ (i.e., divide the covariance of $\tilde{r}_{\text{Sony}}$ with $\tilde{r}_{\text{S\&P500}}$ by the variance of $\tilde{r}_{\text{S\&P500}}$), and the beta of $\tilde{r}_{\text{IBM}}$ with respect to the $\tilde{r}_{\text{S\&P500}}$. Although you should really write $\beta_{\tilde{r}_{\text{IBM}},\tilde{r}_{\text{S\&P500}}}$, no harm is done if you omit the $\tilde{r}$ for convenience, and just elevate the subscripts when there is no risk of confusion. Thus, you can just write $\beta_{\text{IBM,S\&P500}}$, instead.

$$\begin{aligned}
\beta_{\text{IBM,S\&P500}} &= \frac{3.30\%}{(19.0\%)^2} = 0.91 \; ; \\
\beta_{\text{Sony,S\&P500}} &= \frac{4.77\%}{(19.0\%)^2} = 1.31 \\
\beta_{\text{i,j}} &= \frac{Cov(\tilde{r}_{\text{i}}, \tilde{r}_{\text{j}})}{Sdv(\tilde{r}_{\text{i}}) \cdot Sdv(\tilde{r}_{\text{j}})}
\end{aligned} \tag{9.3}$$

Beta is the slope of the best-fitting line when the rate of return on S&P500 is on the $X$-axis and the rate of return on IBM (or Sony) is on the $Y$-axis. Note that although Sony was correlated less with the S&P500 than IBM was correlated with the S&P500, it is Sony that has the steeper slope. Correlation and beta do measure different things. The next chapters will elaborate more on the importance of beta in finance.

The Marquis De Sade would not have been happy. Ok, neither would Mary Poppins.

You have now computed all the statistics that this book will use: means, variances, standard deviations, covariances, correlations, and betas. Only modestly painful, I hope. The next chapter will use no new statistics, but it will show how they work in the context of portfolios.

Solve Now!

**Q 9.3** What is the correlation of a random variable with itself?

**Q 9.4** What is the slope (beta) of a random variable with itself?

**Q 9.5** Return to the historical rates of return on the DAX from Question 9.2 on Page 193. Compute the covariances, correlations and betas for the DAX with respect to each of the other three investment securities.

**Q 9.6** Very advanced Question: Compute the annual rates of return on a portfolio of 1/3 IBM and 2/3 Sony. Then compute the beta of this portfolio with respect to the S&P500. How does this compare to the beta of IBM with respect to the S&P500, and the beta of Sony with respect to the S&P500?

## 9·5 Summary

The chapter covered the following major points:

- Finance often uses statistics based on historical rates of return as standings to predict statistics for future rates of return. This is a leap of faith—often, but not always correct.
- Tildes denote random variables—a distribution of possible future outcomes. In practice, the distribution is often given by historical data.
- The expected rate of return is a measure of the reward. It is often forecast from the historical mean.
- The standard deviation—and its intermediate input, the variance—are measures of the risk. The standard deviation is (practically) the square-root of the average squared deviation from the mean.
- Covariation measures how two variables move together. Causality induces covariation, but not vice-versa. Two variables can covary, even if neither variable would be the cause of the other.
- Like variance, the covariance is difficult to interpret. Thus, covariance is often only an intermediate number on the way to more intuitive statistics.
- The correlation is the covariance divided by the standard deviation of both variables. It measures how reliable a relationship between two variables is. The order of variables does not matter. The correlation is always between $-1$ and $+1$.
- The beta is the covariance divided the standard deviation of the variable on the $X$ axis squared (which is the variance). It measures how steep a relationship between two variables is. The order of variables matters: $\beta_{A,B}$ is different from $\beta_{B,A}$.

# Deeply Dug Appendix

## A More Statistical Theory

### a. Historical and Future Statistics

Physical processes often have known properties. Stock returns do not.

The theory usually assumes that although you do not know the *outcome* of a random variable, you do know the statistics (such as mean and standard deviation) for the outcomes. That is, you can estimate or judge a random variable's unknown mean, standard deviation, covariance (beta), etc. Alas, while this is easy for the throw of a coin or a die, where you know the physical properties of what determines the random outcome, this is not so easy for stock returns. For example, what is the standard deviation of next month's rate of return on PepsiCo?

Use historical statistics as estimators of future statistics.

You just do not have a better alternative than to assume that PepsiCo's returns are manifestations of the same statistical process over time. If you want to know the standard deviation of PepsiCo's next month's rate of return, you typically must assume that each historical monthly rate of return—at least over the last couple of years—was a draw from the same distribution. Therefore, you can use the historical rates of return, assuming each one was an equally likely outcome, to estimate the future standard deviation. Analogously, the mechanics of the computation for obtaining the estimated future standard deviation are exactly the same as those you used to obtain an actual historical standard deviation.

This works well for standard deviations and covariation statistics, but not for means.

But, using historical statistics and then arguing that they are representative of the future is a bit of a leap. Empirical observation has taught us that doing so works well for standard deviations and covariation measures: that is, the historical statistics obtained from monthly rates of returns over the last 3 to 5 years appear to be fairly decent predictors of future betas and standard deviations. Unfortunately, the historical mean rates of return are fairly unreliable predictors of the future rates of returns.

**Q 9.7** When predicting the outcome of a die, why do you not use historical statistics on die throws as predictors of future die throws?

**Q 9.8** Are historical financial securities' mean rates of return good predictors of future mean rates of return?

**Q 9.9** Are historical financial securities' standard deviations and correlations of rates of return good predictors of their future equivalents?

## b. Improving Future Estimates From Historical Estimates

Extreme outcomes have more of two components: higher expected outcome and a higher error term (which will not repeat).

The principal remaining problem in the reliability of historical estimates of covariances for prediction is what statisticians call "regression to the mean." That is, the most extreme historical estimates are likely caused not only by the underlying true estimates, but even more so by chance. For example, if all securities had a true standard deviation of 30% per annum, over a particular year some might show a standard deviation of 40%, while others might show a standard deviation of 20%. Those with the high 40% historical standard deviations are most likely to have lower than their historical standard deviations (dropping back to 30%). Those with the low 20% historical standard deviations are most likely to have higher than their historical standard deviations (increasing back to 30%). This can also manifest itself in market beta estimates. Predicting future betas by running a regression with historical rate of return data is too naïve. The reason is that a stock that happened to have a really high return on one day will show too high a beta if the overall stock market happened to have gone up this day and too low a beta if the overall stock market happened to have gone down this day. Such extreme observations tend not to repeat under the same market conditions in the future.

Shrinkage just reduces the estimate, hoping to adjust for extremes' errors.

Statisticians handle such problems with a technique called "shrinkage." The historical estimates are reduced ("shrunk") towards a more common mean. Naturally, the exact amount by which historical estimates should be shrunk and what number they should be shrunk towards is a very complex technical problem—and doing it well can make millions of dollars. This book is definitely not able to cover this subject appropriately. Still, reading this book, you might wonder if there is something both quick-and-dirty and reasonable that you can do to obtain better estimates of future mean returns, better estimates of future standard deviations, and better estimates of future betas.

Advice: take the average of the market historical statistic and your individual stock historical statistic. It probably predicts better.

The answer is yes. Here is a two minute non-formal heuristic estimation job: To predict a portfolio statistic, average the historical statistic on your particular portfolio with the historical statistic on the overall stock market. There are better and more sophisticated methods, but this averaging is likely to predict better than the historical statistic of the particular portfolio by itself. (With more time and statistical expertise, you could use other information, such as beta, the industry historical rate of return, or the average P/E ratio of the portfolio, to produce even better guesstimates of future portfolio behavior.) For example, the market beta for the overall market is "1.0," so my prescription is to average the estimated beta and 1.0. Commercial vendors of market beta estimates do something similar, too. *Bloomberg* computes the weighted average of the estimated market beta and the number one, with weights of 0.67 and 0.33, respectively, *Value Line* reverses these two weights. *Ibbotson Associates*, however, does something more sophisticated, shrinking beta not towards one, but towards a "peer group" market beta.

An example of shrinking.

↦ *Table 10.7 on Page 235*

Let us apply some shrinking to the statistics in Table 10.7. If you were asked to guesstimate an expected annual rate of return for Wal-Mart over the next year, you would not quote Wal-Marts historical 31.5% as your estimate of Wal-Mart's future rate of return. Instead, you could quote an average of 31.5% and 6.3% (the historical rate of return on the market from 1997 to 2002), or about 20% per annum. (This assumes that you are not permitted to use more sophisticated models, such as the CAPM.) You would also guesstimate Wal-Mart's risk to be the average of 31.1% and 18.7%, or about 25% per year. Finally, you would guesstimate Wal-Mart's market beta to be about 0.95. The specific market index to which you shrink matters little (the Dow-Jones 30 or the S&P500)—but it does matter that you do shrink somehow! An even better target to shrink towards would be the industry average statistics. (Some researchers go as far as to estimate only industry betas, and forego computing the individual firm beta altogether! This is shrinking to a very large degree.) However, good shrinking targets are beyond the scope of this book. Would you like to bet that the historical statistics are better guesstimates than the shrunk statistics? (If so, feel free to invest your money into Wal-Mart, and deceive yourself that you will likely earn a mean return 31.5%! Good luck!)

Here is a summary of some recommendations. Based on regressions using five years of historical monthly data, to predict one-year-ahead statistics, you can use reasonable shrinkage methods for large stocks (e.g., members of the S&P500) as follows:

What works reasonably well, what does not.

| | |
|---|---|
| Mean | Nothing works too well (i.e.,predicting the future from the past). |
| Market-Model Alpha | Nothing works too well. |
| Market-Model Beta | Average the historical beta with the number "1." For example, if the regression coefficient (covariance/variance) is 4, use a beta of 2.5. |
| Standard Deviation | Average the historical standard deviation of the stock and the historical standard deviation of the S&P500. Then increase by 30%, because, historically, for unknown reasons, volatility has been increasing. |

Recall that the market model is the linear regression in which the $x$ variable is the rate of return on the S&P500, and the $y$ variable is the rate of return on the stock in which you are interested.

## c. Other Measures of Spread

There are measures of risk other than the variance and standard deviation, but they are obscure enough to deserve your ignorance (at least until an advanced investments course). One such measure is the mean absolute deviation (MAD) from the mean. For the example of a rate of return of either +25% or −25%,

$$\text{MAD} = 1/2 \cdot |(-25\%)| + 1/2 \cdot |(+25\%)| = 1/2 \cdot 25\% + 1/2 \cdot 25\% = 25\%$$

In this case, the outcome happens to be the same as the standard deviation, but this is not generally the case. The MAD gives less weight than the standard deviation to observations far from the mean. For example, if you had three returns, −50%, −50% and +100%, the mean would be 0%, the standard deviation 70.7%, and the MAD 66.7%.

Another measure of risk is the semivariance ($\mathcal{SV}$), which relies only on observations below zero or below the mean. That is, all positive returns (or deviations from the mean) are simply ignored. For the example of +25% or −25%,

$$\mathcal{SV} = 1/2 \cdot (-25\%)^2 + 1/2 \cdot (0) = 1/2 \cdot 0.0625 = 0.03125$$

The idea is that investors fear only realizations that are negative (or below the mean).

Finally, note that the correlation has another nice interpretation: the correlation squared is the $R^2$ in a bivariate OLS regression with a constant.

## d. Translating Mean and Variance Statistics Into Probabilities

Although you now know enough to compute a measure of risk, you have not bothered to explore how likely outcomes are. For example, if a portfolio's expected rate of return is 12.6% per year, and its standard deviation is 22% per year, what is the probability that you will lose money (earn below 0%)? What is the probability that you will earn 15% or more? 20% or more?

Translating standard deviations into more intuitive risk assessments

It turns out that if the underlying distribution looks like a bell curve—and many common portfolio return distributions have this shape—there is an easy procedure to translate mean and standard deviation into the probability that the return will end being less than $x$. In fact, this probability correspondence is the only advantage that bell shaped distributions provide! Everything else works regardless of the actual shape of the distribution.

Stock returns often assume a normal (bell-shaped) distribution.

For concreteness sake, assume you want to determine the probability that the rate of return on this portfolio is less than +5%:

An example: Z-score and probability.

**Step 1:** Subtract the mean from 5%. In the example, with the expected rate of return of 12.6%, the result is 5% – 12.6% = −7.6%.

**Step 2:** Divide this number by the standard deviation. In ths example, this is −7.6% divided by 22%, which comes to −0.35. This number is called the **Score** or **Z-score**.

**Step 3:** Look up the probability for this Score in the **Cumulative Normal Distribution Table** in Appendix 2·4. For the score of −0.35, this probability is about 0.36.

↦ *Appendix Section 2·4 on Page 781*

In sum, you have determined that if returns are drawn from a distribution with a mean of 12.6% and a standard deviation of 22%, then the probability of observing a single rate of return of +5% or less is about 36%. It also follows that the probability that a return is greater than +5% must be 100% − 36% = 64%.

SIDE NOTE

> In the real world, this works well enough—but not perfectly. Do not get fooled by theoretical pseudo-accuracy. Anything between 30% and 40% is a reasonable prediction here.

How well does the Z-score fit in the history?

Now recall portfolio P in Table 10.1. P had a mean of 12.6% and a standard deviation of 22%. You have just computed that about one third of the 12 annual portfolio returns should be below +5%. 1991, 1993, 1994, 1995, 1996, 1997, 1998, and 1999 performed better than +5%; 1992, 2000, 2001, and 2002 performed worse. As predicted by applying the normal distribution table, about one third of the annual returns were 5% or less.

How likely is it that you will lose money?

A common question is "what is the probability that the return will be negative?" Use the same technique,

**Step 1:** Subtracting the mean of 0% yields 0.0% − 12.6% = −12.6%.

**Step 2:** Dividing −12.6% by the standard deviation of 22% yields the score of −0.57.

**Step 3:** For this score of −0.57, the probability is about 28%.

In words, the probability that the rate of return will be negative is around 25% to 30%. And, therefore, the probability that the return will be positive is around 70% to 75%. The table shows that 4 out of the 12 annual rates of return are negative. This is most likely **sampling error**: with only 12 annual rates of return, it was impossible for the distribution of data to accurately follow a bell shape.

DIG DEEPER

> Many portfolio returns have what is called "fat tails." This means that the probability of extreme outcomes—especially extreme negative outcomes—is often higher than suggested by the normal distribution table. For example, if the mean return were 30% (e.g. for a multi-year return) and the standard deviation were 10%, the score for a value of 0 is −3. The table therefore suggests that the probability of drawing a negative return should be 0.135%, or about once in a thousand periods. Long experience with financial data suggests that this is often much too overconfident for the real world. In some contexts, the true probability of even the most negative possible outcome (−100%) may be as high as 1%, even if the Z-score suggests 0.0001%!

## e. Correlation and Causation

Correlation does not imply Causation.

A warning: covariation is related to, but not the same as **Causation**. If one variable "causes" another, then the two variables will be correlated. But the opposite does not hold. For example, snow and depression are positively correlated, but neither causes the other. Instead, there is another variable (winter) that has an influence on both snow and depression.

Solve Now!

**Q 9.10** If the mean is 20 and the standard deviation is 15, what is the probability that the value will turn out to be less than 0?

**Q 9.11** If the mean is 10 and the standard deviation is 20, what is the probability that the value will turn out to be positive?

**Q 9.12** If the mean is 50 and the standard deviation is 20, what is the probability that the value will turn out to be greater than 80?

**Q 9.13** If the mean is 50 and the standard deviation is 20, what is the probability that the value will turn out to be greater than 30?

**Anecdote: Long Term Capital Management**

**Long-Term Capital Management** (**LTCM**), a prominent hedge fund run by top finance professors and Wall Street traders, collapsed in 1999 in what their quantitative models believed to be a "10 sigma" (i.e., a 10 standard deviation) event. According to the normal distribution probability table, such an event that has a score of −10 should occur with a probability of less than 0.0001%, or 1 in 1,000,000 periods of trading.
You can conclude that either their models were wildly over-optimistic, or their assumption of a normal distribution was incorrect, or the 1 in 1,000,000 actually occurred. Chances are that it was a little bit of all three.
(In essence, LTCM's model believed that it would be exceedingly unlikely that all their bets would go sour at the same time. Of course, they did all go sour together, so the LTCM principals lost most of their wealth.)

[No keyterm list for statistics-g.]

## End of Chapter Problems

## 13 “Solve Now” Answers

1. Do it!
2. The mean was 9.837%. The variance was 7.7%. standard deviation was 27.74%.
3. $$Correlation(\tilde{x},\tilde{x}) = \frac{Cov(\tilde{X},\tilde{X})}{Sdv(\tilde{X})\cdot Sdv(\tilde{X})} = \frac{Var(\tilde{X})}{Var(\tilde{X})} = 1$$
4. $$\beta_{\tilde{x},\tilde{x}} = \frac{Cov(\tilde{X},\tilde{X})}{Var(\tilde{X})} = \frac{Var(\tilde{X})}{Var(\tilde{X})} = 1$$
5. $$Cov(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{S\&P500}}) = 0.0394$$ $$Cov(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{IBM}}) = 0.0464$$ $$Cov(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{Sony}}) = 0.1264$$

   You have already computed the standard deviation of S&P500, IBM, and Sony as 19.0%, 38.8%, and 90.3%; and for the DAX, as 27.74%. Therefore,

   $$Correlation(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{S\&P500}}) = 74.6\%$$ $$Correlation(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{IBM}}) = 43.1\%$$ $$Correlation(\tilde{r}_{\text{DAX}},\tilde{r}_{\text{Sony}}) = 50.5\%$$

   The beta of the DAX with respect to the S&P500 is

   $$\beta_{\tilde{r}_{\text{DAX}},\tilde{r}_{\text{S\&P500}}} = \frac{0.0394}{0.0362} = 1.088$$ $$\beta_{\tilde{r}_{\text{DAX}},\tilde{r}_{\text{IBM}}} = \frac{0.0464}{0.388^2} = 0.308$$ $$\beta_{\tilde{r}_{\text{DAX}},\tilde{r}_{\text{Sony}}} = \frac{0.1264}{0.903^2} = 0.155$$
6. This is to confirm the digging-deeper on Page 215.
7. Because you know the true distribution of future die throws. The historical values are measured with errors.
8. No.
9. Yes, reasonably so.
10. The score is $(0-20)/15 = -1.3$. Therefore, the probability is 9.68%, i.e., roughly 10%.
11. The score is $(0-10)/20 = -0.5$. Therefore, the probability is around 30% that the value will be negative, or 70% that it will be positive.

12. The score is $(80 - 50)/20 = +1.5$. Therefore, the probability is around 93% that the value will be below 80, or 7% that it will be above 80.
13. The score is $(30 - 50)/20 = -1.0$. Therefore, the probability is around 16% that the value will be 30, and 84% that it will be above 30.

**All answers should be treated as suspect. They have only been sketched and have not been checked.**