A First Look at Investments

Historical Rates of Return Background and Market Institutions
The subject of investments is so interesting that I first want to give you a quick tour, instead of laying all the foundation first and showing you the evidence later. I will give you a glimpse into the world of historical returns on the three main asset classes of stocks, bonds, and “cash,” so that you can visualize the main patterns that matter—patterns of risk, reward, and covariation. This chapter also describes a number of important institutions that allow investors to trade equities.

7.1 Stocks, Bonds, and Cash, 1990-2016

Financial investments are often classified into just a few broad asset classes. The three most prominent classes are cash, bonds, and stocks.

Cash: The name cash here is actually a misnomer because it does not designate physical dollar bills under your mattress. Instead, it means debt securities that are very liquid, very low-risk, and very short-term. Other investments that are part of this generic asset class may be certificate of deposits (CDs), savings deposits, or commercial paper. (These are briefly explained in Book Appendix A.) Another common designation for cash is money market. To make our lives easy, we will just join the club and also use the term “cash.”

Bonds: These are debt instruments that have longer maturity than cash. You already know much about bonds and their many different varieties. I find it easiest to think of this class as representing primarily long-term Treasury bonds. You could also broaden this class to include bonds of other varieties, such as corporate bonds, municipal bonds, foreign bonds, or even more exotic debt instruments.

Stocks: Stocks are sometimes all lumped together, and sometimes further categorized into different kinds of stocks. The most common subclassifications for U.S. domestic stocks are as follows:

- The asset class containing a few hundred stocks of the largest firms that trade very frequently is often called large-cap stocks. (Cap is a common abbreviation for “market capitalization,” itself a fancy way of saying “market value.”) Although not exactly true, you can think of the largest 500 firms as roughly the constituents of the popular S&P 500 stock market index. (S&P is Standard and Poor’s. This company invented this index in 1923 and continues to maintain it.) Stocks continuously change in value, disappear, etc. You can very easily invest in an S&P500 basket of stocks by buying a mutual fund or an exchange-traded fund. Our chapter focuses mostly on these large-cap S&P 500 stocks and often just calls them “stocks.”
There are a few thousand other stocks. They are also sometimes put into multiple categories, such as “mid-cap” or “small-cap.” Inevitably, these stocks tend to trade less often, and some seem outright neglected. Small caps can be really small. They may have only $10 million in market cap, and not a single share may be traded for days at a time. In any case, it is so expensive to trade most small-cap stocks that large investors do not bother with them.

There are also other stock-related subclasses, such as industry stock portfolios, or a classification of stocks into “value firms” and “growth firms,” and so on. We shall ignore everything except the large-cap stock portfolio.

Do not take these categories too literally. They may not be representative of all assets that would seem to fit the designation. For example, most long-term bonds in the economy behave like our bond asset class, but some long-term corporate bonds behave more like stocks. Analogously, a particular firm may own a lot of bonds, and its rates of return would look like those on bonds and not like those on stocks. It would also be perfectly reasonable to include more or fewer investments in these three asset classes. (We would hope that such modifications would alter our insights only a little bit.) More importantly, there are also many other important asset classes that we do not even have time to consider, such as real estate, hedge funds, financial derivatives, foreign investments, commodities such as precious metals or orange juice, or art. Nevertheless, cash, bonds, and stocks (or subclasses thereof) are the three most studied financial asset classes, so we will begin our examination of investments by looking at their historical performances.

Graphical Representations of Historical Returns for the S&P 500

Start with Exhibit 7.1. It shows the year-by-year rates of return (with dividends) of the S&P 500. Actually, because of how different sources treat dividends (reinvest or not?), the numbers are never exact. (Some sources even omit dividends in their total rate of return calculation—an exclusion that is definitely wrong.) The series we are using in this book take dividends into account. (And all the numbers are also on the book’s website. Obviously, I do not want to write this textbook with 8 decimal points of precision, so please be aware of—and do not worry about—rounding errors in any of the calculations that follow.) The table and the plot illustrate the same data: You would have lost 3.1% in 1990, gained 30% in 1991, gained 7.4% in 1992, and so on. The average rate of return over the 26 years from 1990-2015 was 10.7% per annum—which I have marked with a dotted line.

Exhibits 7.2 and 7.3 take the same data as in Exhibit 7.1 but present it differently. Exhibit 7.2 shows a histogram that is based on the number of returns that fall within a range. This plot makes it easier to see how spread out returns were—how common it was for the S&P 500 to perform really badly, perform just about okay, or perform really well. For example, the table in Exhibit 7.1 shows that 5 years (2004, 2006, 2010, 2012, and 2014) had rates of return between 10% and 20%. In our 26 years, the most frequent returns were between 0% and 10%. Yet there were also many years that had rates of return below 10%—and even years in which you would have lost more than 20% of your money (such as 1974, 2002, and 2008). And from 2000 to 2002, you would have lost more than a third of your investment! The red triangle indicates that the average rate of return was the aforementioned 10.7%/year.

Most investors are interested in how much money they make and not in statistics. (As Coach Belichick likes to joke, “statistics are for losers.”) Can you take $1 and the 10.7%/average return, and use the compounding formula? Well, this would indicate a final wealth of $1 \cdot 1.107^{26} \approx $14 in 2015. Unfortunately, you would have been far off the mark.

Instead, you need a graph of the compound rate of return, which is shown in Exhibit 7.3. It plots the compounded annual returns (on a logarithmic scale). For example, by the end of 1993, the compound return of $1 invested in 1990 would have been $1.49.

Exhibit 7.1: The Time Series of Rates of Return on the S&P 500 with dividends. The time-series graph is a representation of rates of return of the S&P 500 index (including dividends), as shown in the table above. The average rate of return beginning in 1990 and ending in 2015 was 10.7%/year (indicated by the red triangle and the dotted line); the standard deviation was 17.8%/year. The red box on the right indicates the mean plus or minus the standard deviation.

Original source: CRSP

<table>
<thead>
<tr>
<th>Year</th>
<th>Rate of Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>9.9%</td>
</tr>
<tr>
<td>1991</td>
<td>9.9%</td>
</tr>
<tr>
<td>1992</td>
<td>10.7%</td>
</tr>
<tr>
<td>1993</td>
<td>–3.1%</td>
</tr>
<tr>
<td>1994</td>
<td>10.6%</td>
</tr>
<tr>
<td>1995</td>
<td>–11.9%</td>
</tr>
<tr>
<td>1996</td>
<td>28.4%</td>
</tr>
<tr>
<td>1997</td>
<td>2.1%</td>
</tr>
<tr>
<td>1998</td>
<td>16.0%</td>
</tr>
<tr>
<td>1999</td>
<td>32.5%</td>
</tr>
<tr>
<td>2000</td>
<td>13.5%</td>
</tr>
<tr>
<td>2001</td>
<td>15.0%</td>
</tr>
<tr>
<td>2002</td>
<td>10.7%</td>
</tr>
<tr>
<td>2003</td>
<td>–11.9%</td>
</tr>
<tr>
<td>2004</td>
<td>28.4%</td>
</tr>
<tr>
<td>2005</td>
<td>10.7%</td>
</tr>
<tr>
<td>2006</td>
<td>6.0%</td>
</tr>
<tr>
<td>2007</td>
<td>22.3%</td>
</tr>
<tr>
<td>2008</td>
<td>–4.8%</td>
</tr>
<tr>
<td>2009</td>
<td>20.4%</td>
</tr>
<tr>
<td>2010</td>
<td>–9.0%</td>
</tr>
<tr>
<td>2011</td>
<td>2.1%</td>
</tr>
<tr>
<td>2012</td>
<td>16.0%</td>
</tr>
<tr>
<td>2013</td>
<td>32.5%</td>
</tr>
<tr>
<td>2014</td>
<td>13.5%</td>
</tr>
<tr>
<td>2015</td>
<td>15.0%</td>
</tr>
<tr>
<td>2016</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

There are two further novel aspects to this graph, which is the gray-shaded area. It marks the cumulative CPI inflation. The purchasing power of $1 in 1990 was about the same as $1.87 at the end of 2015. Thus, the $9.82 nominal value in 2015 was really only worth $5.25 in 1990 inflation-adjusted dollars. (And, of course, none of these figures take income taxes into account.)

Many long-term investors make the mistake of compounding the arithmetic average rate of return—commonly just called the mean or average. This would suggest the aforementioned $14 ($1 \cdot (1 + 10.7\%)^{26} - 1$) final wealth. However, if you compound the actual yearly returns, you find that the true compounded investment was only $9.82.

Why are these numbers so different? Think of an example. If you had earned a rate of return of −50% (you lose half) followed by +100% (you double), your compounded rate of return

\[
P_{1/1/1990} \cdot (1 + r_{1990}) \cdot (1 + r_{1991}) \cdot (1 + r_{1992}) \cdot (1 + r_{1993}) = P_{12/31/1993}
\]

The graph also shows inflation.

- Apple.
- Sect. 5.2, Pg. 82.
- Tax Basics.
- Sect. 11.4, Pg. 263.

How to mislead investors: quote arithmetic means for high-volatility investments.
would have been zero. However, your average rate of these two returns would have been a positive \((-50\% + 100\%) / 2 = +25\%\). Equivalently, if you had earned +50\% followed by –50\%, you would have ended up with only 1.5 · 0.5 = 75\% of your investment, a negative rate of return. You will later see a real-world example in which the compound rate of return was –100\% (you lost all your money) but the average rate of return was still positive. Yikes!

### Arithmetic and Geometric Average Rates of Return

The annualized compound rate of return is often called a geometric average. To compute the geometric average, you uncompound (annualize). The annualized rate of return from 1990 to 2015 (26 years) for the S&P 500 investor was

\[
S1 \cdot (1 + r)^{26} \approx $9.82 \quad \iff \quad r \approx \frac{\sqrt[26]{9.82} - 1}{26} \approx 9.2\%
\]

This 9.2\% is about 1.5\% less than the arithmetic rate of return of 10.7\%. The way to interpret this discrepancy is as follows: If there had been no volatility, then a 9.2\% rate of return each and every year would have been enough to compound into $9.82. This can easily lead to misleading comparisons. The historical arithmetic average rate of return for more volatile stocks must be higher than the arithmetic equivalent for less volatile bonds just for you to end up even. If they had the same historical arithmetic rate of return, then bonds would have outperformed stocks.

Unfortunately, the annualized holding rate of return cannot be accurately inferred from the average annual rate of return, and vice-versa. The two are identical only if the rate of return is the same every period (i.e., when there is no risk). Otherwise, the geometric average rate

---

**Exhibit 7.2:** The Statistical Distribution Function of S&P 500 Rates of Return. The graph and table are just different representations of the data in Exhibit 7.1. The x-axis are the individual annual yearly rates of return. The y-axis is the frequency with which these returns occur. Formally, this type of graph is called a density function. It is really just a smoothed version of a histogram.
of return is always less than the arithmetic average rate of return. The more risk, the bigger the difference. It is the geometric rate of return that makes it possible to compare returns with different volatilities in annualized terms. Fortunately, there is an approximation formula.

**Rule of Thumb:** If returns are approximately normally distributed, then the arithmetic mean is higher than the geometric mean by about half the variance.

In our example, the S&P had an annual standard deviation of 17.8%, which comes to a variance of \((0.178)^2 \approx 3.2\%\). Thus, the approximation formula says that the geometric rate of return should have been about 1.6% lower than the arithmetic return. In our case, the approximation is on the money (pun!). A risk-free average rate of return of about 9.2% (which is both geometric and arithmetic) would have allowed you to end up with the same return as the volatile 10.7% arithmetic average rate of return on stocks.

---

**Exhibit 7.3: Compound Rates of Return for the S&P 500.** This graph and table are again just different representations of the same data in Exhibit 7.1. The gray area underneath the figure is the cumulative inflation-caused loss of purchasing power.
Counterintuitive Aspects and Tricks With Quoting

The two averages can be tricky. Let me show you an example.

Say that each period, you can either win or lose 50% (W or L). Your final payoff on a $1 investment is \((1 + r_1) \cdot (1 + r_2) \cdot 1\). For example, if you lose twice, you end up with \((1 - 0.5) \cdot (1 - 0.5) \cdot 1 =$0.25. Thus, your expected payoff \(\mathbb{E}(V)\) is

\[
0.25 \cdot [\text{0.25}] + 0.50 \cdot [0.75] + 0.25 \cdot [2.25] = $1
\]

The probability on the middle term is 50% because it does not matter whether you first win and then lose (WL) or vice-versa (LW). Your average arithmetic rate of return is –50% one quarter of the time (you earned –50% in both periods), 0% half the time (–50% and +50%, one each time), and +50% one quarter of the time. Thus, your expected arithmetic average rate of return \(\mathbb{E}([r_1 + r_2]/2)\) is

\[
0.25 \cdot [0.5] + 0.50 \cdot [0] + 0.25 \cdot [0.5] = 0%
\]

\[\text{Prob}(\text{LL}) \cdot \text{[Mean } r_{1L}] + 2 \cdot \text{Prob}(\text{LW}) \cdot \text{[Mean } r_{1W}] + \text{Prob}(\text{WW}) \cdot \text{[Mean } r_{1W}]\]

And, finally, even your expected two-period compounded rate of return is zero. You earn –75% one quarter of the time, –25% half the time, and +125% one quarter of the time. Thus, \(\mathbb{E}(r_{0,2}) = \mathbb{E}([1 + r_{0,1}] \cdot (1 + r_{1,2}) - 1)\) is

\[
0.25 \cdot [(1 - 0.5) \cdot (1 - 0.5) - 1] + 0.50 \cdot [(1 - 0.5) \cdot (1 + 0.5) - 1] + 0.25 \cdot [(1 + 0.5) \cdot (1 + 0.5) - 1] = 0%
\]

\[\text{Prob}(\text{LL}) \cdot \text{[Compounded } r_{1L}] + 2 \cdot \text{Prob}(\text{LW}) \cdot \text{[Compounded } r_{1W}] + \text{Prob}(\text{WW}) \cdot \text{[Compounded } r_{1W}]\]

This is all just break-even. So far, so good.

However, even though you are breaking even, your geometric average rate of return is less than zero. Your annualized rate of return is \(\sqrt{(1 - 0.5) \cdot (1 - 0.5)} - 1 = -50%\) one quarter of the time, \(\sqrt{(1 - 0.5) \cdot (1 + 0.5)} - 1 = -13.4\%\) half the time, and \(\sqrt{(1 + 0.5) \cdot (1 + 0.5)} - 1 = +50%\) one quarter of the time. Your expected geometric rate of return is therefore

\[
0.25 \cdot [\sqrt{(1 - 0.5) \cdot (1 - 0.5)} - 1] + 0.50 \cdot [\sqrt{(1 - 0.5) \cdot (1 + 0.5)} - 1] + 0.25 \cdot [\sqrt{(1 + 0.5) \cdot (1 + 0.5)} - 1] \approx -6.7%
\]

\[\text{Prob}(\text{LL}) \cdot \text{Annualized } r_{1L} + 2 \cdot \text{Prob}(\text{LH}) \cdot \text{Annualized } r_{1H} + \text{Prob}(\text{LL}) \cdot \text{Annualized } r_{1H}\]

The square-root is “at fault” here. It is why this expected geometric average rate of return is negative. You can interpret this geometric average as stating that a negative geometric rate of return of –6.7% would have been enough to keep your true expected payoff at your original investment level of $1 (because earning a compounding 50% is really great!).

But, if you had exactly one-half of the time a rate of return of –50% and one-half of the time a rate of return of +50%, over time, you would lose money; because \((1 - 0.5) \cdot (1 + 0.5) \approx -13.4\%\). The difference between the –13.4% (well, half the time) and the 0% is the difference between simultaneously-equally-likely realizations and sequential realizations. If you get either –50% or +50% with equal probability, your expected value is 0. If you get –50% followed by +50%, your expected value is negative.

Are You Expecting Compound Arithmetic or Geometric Historical Averages?

Unfortunately, this is not just an academic egghead concern. If you knew the population distribution, the distinction between arithmetic and geometric returns would be just a footnote. However, you usually do not. The conceptual problem now is that the “statistical sampling” logic implicitly converts the historical sequential realizations into assumed simultaneous draws in each period. Here is what I mean. Let’s say that you had observed just two periods in which investors had first earned one -50% rate of return and then one +50% rate of return. A $1 investment left them with $0.75 over this two-year period. This is all you know. If you now assume that you will receive either -50% or +50% with equal probability in each of the following two periods ($0.50 or $1.50 after one year; $0.25, $0.75, $0.75, $2.25 after two years), then you will expect to end up with $1. If you use historical realizations as equally likely samples, you are guessimating that you will do better in the future ($1) than in the past ($0.75), purely based on the past. One way around this problem is to work with compound rates of return to begin with. Over two years, you earned -25%. This is all you know. Thus, if you have to guess what you will earn over the next two years, it is also -25%.

Let’s put this insight to work on the specific question at hand. Can you estimate how much $1 invested in the S&P 500 will be worth in 43 years, given the data? In the data, the arithmetic average rate of return was 11.3% per year. If you knew the rate of return had a true population mean of 11.3% each and every year in the future (and you had no uncertainty about the population mean), then you would expect to earn $0.75 per year. Thus, you should expect to end up with $1 in 43 years. If you use historical realizations as equally likely samples, you are guessimating that you will do better in the future ($1) than in the past ($0.75), purely based on the past. Which way is the most appealing? After all, investors received only $9.82 over the last 26 years. Would you not expect your next 26-year performance to be $9.82, too? (This would suggest you compound 9.2%, not 10.7%.) This is what you would guess it to be if you assumed that the last 26 years were just one grand realization, not 43 individual realizations.

The long-standing convention in most NPV applications, where you often have to estimate an equivalent rate of return in the stock market as your opportunity cost of capital (in the PV denominator), is to compound annual or even monthly historical arithmetic rate of returns—leading NPV users to expect $14 as an opportunity cost in this case. However, it is not at all clear that this is correct. One can argue that the historical return of $10 (or $9.82 if you want to misleadingly pretend that we have this kind of accuracy) gives you a better estimate, that the historical arithmetic rate of return compounded to $14 gives you a better estimate, that something in between $10 and $14 gives you a better estimate—and, most counterintuitively (reasoning omitted), even that a value above $14 gives you a better estimate. You have been warned! In any case, don’t forget your basics: the problem is your estimation uncertainty. Your goal is estimating what alternative investments would earn elsewhere compared to your own project. Statistics and math are only aids, not gospels.

Q 7.1. What can you see in a time-series graph that is not in a histogram?
Q 7.2. What can you see in a histogram that is more difficult to see in a time-series graph?
Q 7.3. What can you see in a compound return graph that is not in the time-series graph?
Q 7.4. What is the annualized holding rate of return and the average rate of return for each of the following?

1. An asset that returns 5% each year.
2. An asset that returns 0% and 10% in alternate years.
3. An asset that returns -10% and 20% in alternate years.

Is the distance between the two returns larger when there is more risk?
Q 7.5. If the risk-free rate of return is 4% per annum, how big is the difference between the arithmetic and the geometric average rate of return?
Historical Performance for a Number of Investments

Stocks, Bonds, and Cash

What does history tell you about rate of return patterns on the three major investment categories—stocks, bonds, and cash? You can find out by plotting exactly the same graphs as those in Exhibits 7.1, 7.2, and 7.3. Exhibit 7.4 repeats them for cash, bonds, and stocks all on the same scale. You have already seen the third row, but I have changed the scale to make it easier to make direct comparisons to the other two asset classes. These mini-graphs display a lot of information about the performance of these investments.

<table>
<thead>
<tr>
<th>Cash:</th>
<th>Mean (Reward): 3.0%/yr.</th>
<th>Std.Dev (Risk): 2.4%/yr.</th>
<th>$1 in 1/1990 would have become $2.14 in 12/2015</th>
<th>Correl w/ S&amp;P: 9%</th>
<th>Beta w/ S&amp;P: 0.01</th>
</tr>
</thead>
</table>

| Bonds: | Mean (Reward): 8.3%/yr. | Std.Dev (Risk): 12.6%/yr. | $1 in 1/1990 would have become $7.64 in 12/2015 | Correl w/ S&P: -24% | Beta w/ S&P: -0.17 |

| Stocks: | Mean (Reward): 10.7%/yr. | Std.Dev (Risk): 18.0%/yr. | $1 in 1/1990 would have become $9.82 in 12/2015 | Correl w/ S&P: +100% | Beta w/ S&P: +1.0 |

*The first three rows show historical returns for the three asset classes.*

So let’s compare the first three rows:

**Cash** in the first row is the overnight Federal Funds interest rate. Note how tight the distribution of cash returns was around its 3% mean. You would never have lost money (in nominal terms), but you would rarely have earned much more than its mean. The value of your
total investment portfolio would have steadily marched upward—although pretty slowly. Each dollar invested on January 1, 1990 would have become $2.14 at the end of 2015. Of course, inflation would have eroded the value of each dollar. In purchasing power, your $1 would have been the equivalent of $2.14/1.87 ≈ $1.14 in 2015—and this is before the inevitable taxes you would have had to pay.

**Bonds** in the second row are long-term Treasury bonds. The middle graph shows that the bars are now sometimes slightly negative (years in which you would have earned a negative rate of return)—but there are now also years in which you would have done much better than cash. This is why the histogram is much wider for bonds than it is for cash: Bonds were riskier than cash. The standard deviation tells you that bond risk was 12.6% per year, much higher than the 2.4% cash risk. Fortunately, in exchange for carrying more risk, you would have also enjoyed an average rate of return of 8.1% per year, which is a lot higher than the 3.0% of cash. And $1 invested in 1990 would have become not just $2.14 but $7.64 ($4.09 in real terms)—again before taxes.

**Stocks** in the third row are our familiar portfolio of S&P 500 firms. Annual rates of return here are with dividends, and thus always more than the percent change in the widely quoted S&P500 index. The left graph shows that large stocks would have been even riskier than bonds. The stock histogram is more "spread out" than the bond histogram. The middle graph shows that there were years in which the negatives of stocks could be quite a bit worse than those for bonds, but that there were also many years that were outright terrific. And again, the higher risk of stocks also came with more reward. The S&P 500’s risk of 18% per year was compensated with a mean rate of return of 10.7% per year. Your $1 invested in 1990 would have ended up being worth $9.82 in 2015 ($5.25 in real terms)—again before taxes (although taxes are usually a little lower on stocks than on bonds).

The difference between $9.82 in stocks and $2.14 in cash or $7.64 in bonds is an understatement if you are a common taxable retail investor. Nominal interest payments would have been taxed each year at your full income-tax rate, somewhere between 30% and 50% per year. In contrast, the capital gains on stocks would have been taxed only at the end and at the much lower capital gains tax rate, between 15% and 30%. Roughly speaking, taking taxes into account, if you had invested in cash, you would have ended up with less real purchasing power than you started with. You would have gained real purchasing power in bonds (maybe $2.00). And you would have roughly quadrupled your purchasing power in stocks. The sample shows good years for stocks—and perhaps even unusually good. Not every historical 43-year period would have shown this large a difference between cash and stocks. The difference between bonds and stocks were much more modest, but still considerable if you had to take income taxes into account.

**More Asset Classes**

Exhibit 7.5 shows the performance of a few other large asset classes and over a longer time period (though not ending in 2015). Small-firm stocks were riskier (and more difficult to trade), but their average rates of return were higher. Corporate bonds sat between government bonds and stocks in terms of reward, although their risk was comparable to the former. Intermediate government bonds (i.e., with about 5-year maturity) were somewhere between cash and long-term bonds. Gold was an extremely risky investment by itself, but it also did well over the sample. (Not shown in this table, it did well in years when stocks did poorly.) Moreover, unlike bonds, gold’s gains were taxed at the lower capital-gains rate. Housing is the average price appreciation of residential houses. It probably understates the rate of return by about 3-6% per year, because it omits the value and other costs of living in a house. Owning real-estate (a house) from 1970 to 2010 was a good investment, especially if you take into account that tax rules now shelter some gains from any taxes. However, the 6-7% risk is misleading. Many economists believe that there was a housing bubble in the 2000s, which explains both the fantastic appreciation and the
### Other Samples

<table>
<thead>
<tr>
<th>Sample Description</th>
<th>Time Period</th>
<th>Annualized Return (Geo, Ari)</th>
<th>SDv (95% Confidence Interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-U.S. OECD Equities vs. Bonds</td>
<td>1900-2010</td>
<td>≈ 3.8% (5.0% ari, SDv ≈ 15.5%)</td>
<td></td>
</tr>
<tr>
<td>U.S. Equities vs. Bonds</td>
<td>1900-2010</td>
<td>≈ 4.4% (6.4% ari, SDv ≈ 20.5%)</td>
<td></td>
</tr>
<tr>
<td>Buyout Funds</td>
<td>1984-2008</td>
<td>14% (ari), ≈ 1.2 × S&amp;P500 (geo cum)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000-2008</td>
<td>10% (ari), ≈ 1.3 × S&amp;P500 (geo cum)</td>
<td></td>
</tr>
<tr>
<td>VC Funds</td>
<td>1984-2008</td>
<td>17% (ari), ≈ 1.4 × S&amp;P500 (geo cum)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2000-2008</td>
<td>−1% (ari), ≈ 0.9 × S&amp;P500 (geo cum)</td>
<td></td>
</tr>
<tr>
<td>Art</td>
<td>1976-2004</td>
<td>≈ 6% (SDv ≈ 9%) vs. Stocks 12% (15%)</td>
<td></td>
</tr>
<tr>
<td>Wine (not consumed)</td>
<td>1996-2001</td>
<td>≈ 20% (SDv ≈ 8%)</td>
<td></td>
</tr>
<tr>
<td>Commodities Futures</td>
<td>1959-2004</td>
<td>≈ 10% (SDv ≈ 12%) vs. Stocks 6% (15%)</td>
<td></td>
</tr>
<tr>
<td>... Spot</td>
<td>1959-2004</td>
<td>≈ 4%</td>
<td></td>
</tr>
</tbody>
</table>

### Exhibit 7.5: Comparative Investment Performance for More Asset Classes and Samples. The upper panel was calculated by Your’s Truly. Original data sources, see leftmost column: L= London Gold Exchange, I=Ibbotson Stocks, Bonds, Bills and Inflation, SBDI Valuation Yearbook, Morningstar 2011. S= Robert Shiller, Irrational Exuberance, 2nd Ed., US National Index. Note that housing appreciation ignores the useful housing rental yield, and thus understates the rate of return. The lower panel has a potpourri of quotes from different papers with different samples and methods. The buyout funds and VC funds geometric performance are quoted over the entire 25-year sample. The commodities include metals, agricultural, and energy, but not financials.

subsequent crash. From 1992 to 2006, there was not a single year in which prices declined. But from 2007 to 2009, residential houses lost about 30% of their values. (Real estate investment trusts [REITs] are another interesting way to invest in real-estate.) Unfortunately, I did not have the data to replicate these calculations for some further asset classes in the lower panel of Exhibit 7.5. Because they are over different intervals, I have tried to include the same-period performance of stocks. They are not only “fun” to look at, but worth contemplating as potential investments, too.

Individual Stocks

Instead of buying entire asset classes, you could also have bought just an individual stock. How would such holdings have differed from an investment in the broader asset class “stocks”? Exhibit 7.6 keeps the same scale but now shows the rates of return of a few sample stalwart firms: Coca-Cola [KO], PepsiCo [PEP], Intel [INTC], and United Airlines [UAL]. For comparison, the bottom is again the S&P 500. You can see that individual stocks’ histograms are really wide: Investing in a single stock would have been a rather risky venture, even for these four household names. Indeed, it is not even possible to plot the final year for UAL in the rightmost compound return graph, because UAL stock investors lost all invested money in the 2003 bankruptcy, which on the logarithmic scale would have been minus infinity. And UAL illustrates another important issue: Despite losing all the money, it still had a reasonable average rate of return. If you extended the sample backwards a little, it would be positive, even though you would have still lost all your money assuming reinvestment of dividends into the stock. (You already know why: This was the difference between geometric and arithmetic averages explained on Page 143.)

Q 7.6. Rank the following asset categories in terms of risk and reward: cash (money market), long-term bonds, the stock market, and a typical individual stock.

Q 7.7. Is the average individual stock safer or riskier than the stock market?

Q 7.8. Is it possible for an investment to have a positive average rate of return, but still lose you every penny?

Comovement, Market Beta, and Correlation

Exhibit 7.7 highlights the rates of return on the S&P 500 and one specific stock, Intel. The top rows redraw the time-series graphs for these two investments. Do you notice a correlation between these two series of rates of return? Are the years in which one is positive (or above its mean) more likely also to see the other be positive (or above its mean), and vice-versa? It does seem that way. For example, the worst rates of return for both were 2002 and 2008—and even more so for Intel investors than market investors. In contrast, 1992, 1995, and 1998 were good years for both. And again, even more so for Intel investors. The correlation is not perfect: In 2004, the S&P 500 had a good year, but Intel had a bad one; and in 2001, the market had a bad year, but Intel turned out alright. It is very common for all sorts of investments in the economy to move together with the stock market: In years of malaise, almost all assets tend to be in malaise. In years of exuberance, almost all tend to be exuberant. This tendency is called comovement.

The comovement of investments is very important if you do not like risk. An investment that increases in value whenever the rest of your portfolio decreases in value is practically like “insurance” that pays off when you need it most. You might buy into such an investment even if it offers only a very low expected rate of return. In contrast, you might not like an investment that does very badly whenever the rest of your portfolio also does badly. To be included in your portfolio, such an investment would have to offer a very high expected rate of return.
### Exhibit 7.6: Comparative Investment Performance, 1990-2015. Data source: CRSP

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean (Reward)</th>
<th>Std.Dev (Risk)</th>
<th>$1 in 1990 would have become $100,000 in 2015</th>
<th>Correl w/ S&amp;P</th>
<th>Beta w/ S&amp;P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Coca-Cola (KO):</strong></td>
<td>13.0%/yr</td>
<td>21.9%/yr</td>
<td>$15.22 in 2015</td>
<td>61%</td>
<td>0.8</td>
</tr>
<tr>
<td><strong>PepsiCo (PEP):</strong></td>
<td>12.9%/yr</td>
<td>17.9%/yr</td>
<td>$16.93 in 2015</td>
<td>55%</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Intel (INTC):</strong></td>
<td>23.4%/yr</td>
<td>44.0%/yr</td>
<td>$45.42 in 2015</td>
<td>63%</td>
<td>1.6</td>
</tr>
<tr>
<td><strong>United (UAL):</strong></td>
<td>-9.4%/yr</td>
<td>57.7%/yr</td>
<td>$0.00 in 2015</td>
<td>64%</td>
<td>1.9</td>
</tr>
<tr>
<td><strong>Stocks:</strong></td>
<td>10.7%/yr</td>
<td>17.8%/yr</td>
<td>$9.82 in 2015</td>
<td>100%</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Exhibit 7.7: Rates of Return on the S&P 500 and Intel (INTC). The top left graph plots the annual rates of return on the S&P 500; the top right graph plots the annual rates of return on Intel. The bottom left graph combines the information. The stock market rate of return is on the x-axis, the Intel rate of return is on the y-axis. The figure shows that in years when the stock market did well, Intel tended to do well, too, and vice-versa. This can be seen in the slope of the best-fitting line (in fat blue), which is called the market beta of Intel. The market beta will play an important role in later investments chapters. On the bottom left, the slope is about 1.3 (steeper than 45 degrees). On the bottom right, I used 3 years of daily returns. The slope is about 1.0. Trust me when I tell you that the daily-based line on the right turns out to be a better estimator of future market-betas than the annual-based line on the left—especially if the historical slope is averaged with 1.0.
How can you measure the extent to which securities covary with others? For example, how did Intel covary with the S&P 500 (our stand-in for the market portfolio)? Did Intel also go down when the market did (making a bad situation worse), or did it go up (thereby serving as useful insurance)? How can you quantify such comovement?

You can answer this graphically. Plot the two return series against one another, as in the bottom plots in Exhibit 7.7. Then find the line that best fits between the two series. (You will learn later how to compute such “regression” lines.) The slope of this line is called the market beta of a stock, and it is a measure of comovement between the rate of return on the stock with the rate of return on the market. It tells an investor whether this stock moved with or against the market. It carries great importance in financial economics.

- If the best-fitting line has a slope that is steeper than the 45° diagonal (well, if the x- and y-axes are drawn with the same scale), then the market beta is greater than 1. Such a line would imply that when the stock market did better (the x-axis), on average your stock did a lot better (the y-axis). For example, if a stock has a very steep positive slope—say, +3—then (assuming you hold the market portfolio) if the market dropped by an additional 10%, this stock would have been expected to drop by an additional 30%. If you primarily held the market portfolio, this new stock would have made your bad situation worse.
- If the slope is less than 1 (or even 0, a plain horizontal line), it means that, on average, your stock did not move as much (or not at all) with the stock market.
- If a stock has a very negative slope such as −2, this investment would likely have “rescued” you when the market dropped by 10%. On average, it would have earned a positive 20% rate of return. Adding such a stock to your market portfolio would be like buying insurance.

Intel’s annual rates of return had a slope of 1.6 against the market. That is, it was steeper than the diagonal line. In effect, this means that if you had held the stock market, Intel would have been an additional hazard for you. A 1% performance above (below) normal for the S&P would have meant you would have expected to earn 1.6% above (below) normal in your Intel holdings. However, for estimating the future market-beta in the real world, it turns out that 3 years of recent daily stock returns is better practice. (After running the statistical procedure called regression analysis to obtain the best line fit, for forecasting the future beta, you should also take the average of your regression beta estimate and 1.0.) As of 2015, it turns out that Intel’s best forward-looking market-beta is much lower. In fact, it is just about 1.

Instead of beta, you could measure comovement with another statistic that you may already have come across: the so-called correlation. Correlation and beta are related. The correlation has a feature that beta does not. A correlation of +100% indicates that two variables always perfectly move together; a correlation of 0% indicates that two variables move about independently; and a correlation of −100% indicates that two variables always perfectly move in opposite directions. (A correlation can never exceed +100% or −100%.) In Intel’s case, one can work out that the correlation is +63%. The correlation’s limited range from −1 to +1 is both an advantage and a disadvantage. On the positive side, the correlation is a number that is often easier to judge than beta. On the negative side, the correlation has no concept of scale. It can be 100% even if the y variable moves only very, very mildly with x (e.g., if every y = 0.0001 · x, the correlation is still a positive 100%). In contrast, beta can be anything from minus infinity to plus infinity.

A positive correlation always implies a positive beta, and vice-versa. Of course, beta and correlation are only measures of average comovement: Even for investments with positive betas, there are individual years in which the investment and stock market do not move together (e.g., 2004 for Intel and the S&P 500). Stocks with negative betas, for which a negative market rate of return on average associates with a positive stock return (and vice-versa), are rare. There are only a very few investment categories that are generally thought to be negatively correlated with the market—principally gold and other precious metals. Interestingly, long-term Treasury bonds
nowadays seem to have negative correlation with the market. It used to be the opposite. I am not confident to tell you what it will be in the future.

**Q 7.9.** How do you graph a “market beta”? What should be on the x-axis, and what should be on the y-axis? What is an individual data point?

**Q 7.10.** What is the market beta of the market?

---

### Causality vs. Correlation

Allow me a brief diversion. The most important problems in finance—and economics, and statistics, and science, and theory, and practice—may well be whether correlation implies causation. If X causes Y, the two should be correlated. The problem is that if X does not cause Y, the two can still be correlated. People do not have very good intuition about the distinction. Thus, they often commit serious and harmful interpretation mistakes. (Some are deliberate, as in political demagoguery.) For example, just because a stock had tended to go up when there was sunspot activity (or when the overall stock market went up) does not mean that you have found that sunspots are one cause of stocks going up—this is an example of spurious correlation. (In fact, over long sample periods, increased sunspot activity has indeed been associated with higher stock prices.) Just because good CEOs are paid high salaries average does not mean that paying more money would make your own CEO any better—the causation may go the other way. And just because increases in government spending are associated with reductions in unemployment does not mean that the government can reduce unemployment by spending more—some other economic factor may have determined both spending and unemployment. Determining causality is important if you want to know how your strategies and policies are likely to change outcomes.

An early answer to measuring causality in economics came from two econometricians: if unexpected changes in X predict unexpected changes in Y, then X may cause Y. If you saw an unusual sunspot (X) and it usually preceded (in time) an unusual increase in the stock return (Y), then it is a hint that X causes Y. This concept is called Granger-Sims causality. By this metric, the data reject the hypothesis that sunspots have “caused” stock returns to go up. (Both sunspots and stock prices happened to go up; it was time effects that induced the spurious correlation.) Unfortunately, Granger-Sims causality isn’t perfect, either. By its metric, the weather forecast “causes” the weather. (Unusual changes in the forecast indeed predict subsequent unusual changes in weather!)

Just when we were ready to give up, economics stumbled upon an approach that is now called “quasi-experimental.” It is revolutionizing empirical economics right now (and soon economic consulting, too). Let me illustrate this with an example. Think about figuring out whether access to loans increases the success of startups. The problem is that ventures that are less likely to be successful are also less likely to attract lenders. Thus, it is not possible to conclude from the fact that funded ventures had higher success rates in the past (which they did!) that loan access played a critical role in this success. Funding may have been more like the weather forecast, itself responding to other factors (e.g., promising business plans) that ultimately determined project success. If loans to startup did not have a positive influence on survival, government programs that seek to make more loans available to more startups would probably not be a good idea.

But Fracassi, Garmaise, Kogan, and Natividad have an answer! It turns out that a particular lender employed an automated credit score algorithm with a cutoff that determined loan funding. Applicants with scores just above the cutoff (say, 4.14) received a loan. Applicants with scores below (say, 4.13) were denied. The probability of survival was 30% for the 4.14 group and 25% for the 4.13 group. Because these applicants were so close in score, their differences were

---

**Caused on Correlation:**


*Please note, the provided URL is a placeholder and the actual content might be different.*
Diversification, Section B.2, Pg.171.

probably just noise. Thus, it is likely that the entire 5% was due to the fact that the 4.14 firm got the loan and the 4.13 firm did not. (To make sure, they also compared their 5% survival difference to the survival differences for firms between 4.12 and 4.13 and firms between 4.14 and 4.15, neither of which was treated differently by the lender and neither of which showed any differences in survival.) This is convincing evidence that access to loans indeed helped improve the chances of survival for the startup firms in their sample.

The Big Picture Take-Aways

What can you learn from the performance and correlation graphs? Actually, almost everything there is to learn about investments! I will explain these facts in much more detail soon. In the meantime, here are the most important points that the graphs show:

• History tells us that stocks have offered higher average rates of return than bonds, which in turn have offered higher average rates of return than cash. However, keep in mind that this was only on average. In any given year, the relationship might have been reversed. For example, stocks lost 22% of their investment in 2002, while cash gained about 1.7%.

• Although stocks did well (on average), you could have lost your shirt investing in them, especially if you had bet on just one individual stock. For example, if you had invested in United Airlines in 1990, you would have lost all your money.

• Cash was the safest investment—its distribution is tightly centered around its mean, so there were no years with negative returns. Bonds were riskier. And stocks were even riskier. (Sometimes, stocks are said to be “noisy,” because it is really difficult to predict how they will perform.)

• There seems to be a relationship between risk and reward: Riskier investments tended to have higher average rates of return. (However, you will learn soon that risk has to be looked at in context. Thus, please do not overread the simple relationship between the mean and the standard deviation here.)

• Large portfolios consisting of many stocks tended to have less risk than individual stocks. The S&P 500 stocks had a risk of around 15-20% per year, which was less than the risk of most individual stocks. (For example, even large-firm stalwarts like Coca-Cola, PepsiCo, and Intel had risks between 18% and 60%). This is due to “diversification,” a concept we will discuss in the next chapter.

• The average rate of return is always larger than the geometric (compound) rate of return. A positive average rate of return usually, but not always, translates into a positive compound holding rate of return. For example, if investment in United Airlines had started in 1970, it would have had a positive average rate of return, despite having lost all its investors’ money.

• Stocks tend to move together. For example, if you look at 2001 and 2002, not only did the S&P 500 go down, but most individual stocks tended to go down, too. In 1998, on the other hand, most stocks tended to go up (or at least not down much). The mid-1990s were good to stocks. In contrast, money-market returns had little to do with the stock market.

• On an annual frequency, the correlation between the stock market (the S&P 500) and cash was small (about 10%). The correlation between the stock market and bonds used to be positive, but now seems to be negative. The correlation between individual stocks and the stock market was around 50% to 70%. The fact that investment rates of return tend to move together is important. It is the foundation for the market beta, a measure of risk that we have touched on and that will be explained in detail in Chapter 8.

Will History Repeat Itself?

As a financier, you are not interested in history for its own sake. Instead, you really want to know more about the future. History is useful only because it is your best available indicator of the future. But which history? One year? Thirty years? One hundred years? I can tell you that if you had drawn the graphs from this chapter beginning in 1926 instead of in 1970, the big conclusions would have remained the same. However, if you had started in 2001, it would have looked differently. What would you have seen? Four awful years for stock investors in the sample. You should feel intuitively that this might not be a good representative sample period. To make any sensible inferences about what is going on in the financial markets, you need many years of history, not just a decade or so—and certainly not the 6-week investment performance touted by some funds or friends (who also often display remarkable selective memory!). The flip side of this argument is that you cannot reliably say what the rate of return will be over your next year. It is easier to forecast the average annual rate of return over five to ten years than over one year. Your investment outcome over any single year will be very noisy.

Instead of relying on just one year, relying on statistics computed over many years is much better. However, although twenty to thirty years of performance is the minimum number necessary to learn something about return patterns, this is still not sufficient for you to be very confident. Again, you are really interested in what will happen in the next five to ten years, not what happened in the last five to ten years. Yes, the historical performance can help you judge but you should not trust it blindly. For example, an investor in UAL in 2000 might have guessed that the average rate of return for UAL would have been positive—and would have been sorely disappointed. Investors in the Japanese stock market in 1986 saw the Nikkei-225 stock market index rise from 10,000 to 40,000 by 1990—a 40% rate of return per year. If they had believed that history was a good guide, they would have expected 40,000 \( \cdot 1.40^{13} \approx 3.2 \) million by the end of 2002. Instead, the Nikkei fell below 8,000 in April 2003 and has only recently recovered to 19,000 as of early 2017. History would have been a terrible guide.

Nevertheless, despite the intrinsic hazards in using historical information to forecast future returns, having historical data is a great advantage. It is a rich source of forecasting power, so like everyone else, you will have to use historical statistics. But please be careful not to rely too much on them. For example, if you look at an investment that had extremely high or low past historical rates of return, you may want to think that similar performance is likely to continue.

In relative terms, what historical information can you trust more and what historical information should you trust less?

**Historical risk:** Standard deviations and correlations (how stock movements tend to be related or unrelated) tend to be fairly stable, especially for large asset classes and diversified portfolios. That is, say, for 2015 to 2020, you can reasonably expect PepsiCo to have a risk of about 25-30% per year, a correlation of about 50-70% with the market, and a market beta of about 0.7 to 1.1. The estimates slowly deteriorate in accuracy over multi-year horizons, though.

**Historical mean reward:** Historical average rates of return are not very reliable predictors of future expected rates of return. That is, you should not necessarily believe that PepsiCo will continue to earn an expected rate of return of 13% per year. And, for sure, you should not expect Intel to expect to earn 23.4% in the future.

**Realizations:** You should definitely not believe that past realizations are good predictors of future single-year realizations. Just because the S&P 500 had earned about 11% on average does not make it particularly likely that it will have a rate of return of 11% in any one given future year.

A lottery analogy may help you understand the last two points better. If you have played the lottery many times, your historical average rate of return is unlikely to be predictive of your
future expected rate of return—especially if you have won it big at least once. Yes, you could
trust it if you had millions of historical realizations, but you inevitably do not have so many.
Consequently, your average historical payoff is only a mediocre predictor of your next week’s
payoff. And you should definitely not trust your most recent realization(s) to be indicative of the
future. Just because “5, 10, 12, 33, 34, 38” won last week does not mean that it will likely win
again.

Henceforth, like almost all of finance, we will just assume we know the statistical distributions
from which future investment returns will be drawn. For exposition, this makes our task a
lot easier. When you want to use our techniques in the real world, you will usually collect
historical data and pretend that the future distribution is the same as the historical distribution.
(Some investors in the real world use some more sophisticated techniques, but ultimately these
techniques are also just variations on this theme.) However, always remember: historical data is
an imperfect guide to the future.

7.2 Overview of Equity-Related Market Institutions

Let’s look into the institutional arrangements for equity trading. After all, from our corporate
perspective, stocks are more interesting than many other financial instruments, such as foreign
government bonds, even if there is more money in foreign government bonds than in corporate
equity. After all, it is the equity holders who finance most of the risks of corporate projects.
Moreover, although there is more money in nonequity financial markets, the subject area of
investments often focuses on equities (stocks), too, because retail investors find it easier to buy
stocks, and historical data for stocks is relatively easy to come by. So it makes sense to describe a
few institutional details as to how investors and stocks “connect”—exchange cash for claims,
and vice-versa.

Brokers

Most individuals place their orders to buy or sell stocks with a retail broker, such as Ameritrade
(a “deep-discount” broker), Charles Schwab (a “discount” broker), or Merrill Lynch (a “full-service
broker). Discount brokers may charge only about $10 commission per trade, but they often
receive “rebate” payments back from the market maker to which they route your order. This
is called “payment for order flow.” The market maker in turn recoups this payment to the
broker by executing your trade at a price that is less favorable. Although the purpose of such an
arrangement seems deceptive, the evidence suggests that discount brokers are still often cheaper
in facilitating investor trades—especially small investor trades—even after taking this hidden
payment into account. They just are not as (relatively) cheap as they want to make you believe.

Investors can place either market orders, which ask for execution at the current price, or limit
orders, which ask for execution if the price is above or below a limit that the investor can specify.
(There are also many other modifications of orders, e.g., stop-loss orders [which instruct a broker
to sell a security if it has lost a certain amount of money], good-til-canceled orders, and fill-or-kill
orders.) The first function of retail brokers, then, is to handle the execution of trades. They
usually do so by routing investors’ orders to a centralized trading location (e.g., a particular stock
exchange), the choice of which is typically at the retail broker’s discretion, as is the particular
individual (e.g., floor broker) engaged to execute the trade. The second function of retail brokers is
to keep track of investors’ holdings, to facilitate buying on margin (whereby investors can borrow
money to buy stock, allowing them to buy more securities than they could afford on a
pure cash basis), and to facilitate selling securities “short,” which allows investors to speculate
that a stock will go down.

Many large institutional investors separate the two functions: The investor employs its own
traders, while the broker takes care only of the bookkeeping of the investor’s portfolio, margin
provisions, and shorting provisions. Such limited brokers are called prime brokers.
How Shorting Stocks Works

If you want to speculate that a stock will go down, you would want to short it. This shorting would be arranged by your broker. Shorting is important enough to deserve an extended explanation:

- You find an investor in the market who is willing to lend you the shares. In a perfect market, this does not cost a penny. In the real world, the broker has to find a willing lender. Both the broker and lender usually earn a few basis points per year for doing you the favor of facilitating your short sale.
- After you have borrowed the shares, you sell them into the market to someone else who wanted to buy the shares. In a perfect market, you would keep the proceeds and earn interest on them. In the real world, your broker may force you to put these proceeds into low-yield safe bonds. If you are a small retail investor, your brokerage firm may even keep the interest proceeds altogether.
- When you want to “unwind” your short, you repurchase the shares and return them to your lender.

For example, if you borrowed the shares when they were trading for $50 (and sold them into the market), and the shares now sell for $30, you can repurchase them for $20 less than what you sold them for into the market. This $20 is your profit. In an ideal world, you can think of your role effectively as the same as that of the company—you can issue shares and use the $50 proceeds to fund your investments (e.g., to earn interest). In the real world, you have to take transaction costs into account. (Shorting has become so common that there are now exchange-traded futures on stocks that make it even easier. Shorting is also common and easy for bonds.)

Shorting the S&P 500 or some other market indices is even easier than shorting individual stocks. You can either short the relevant index ETF (explained below), which works the same way as shorting any other stock); or you can sell traded Futures on common stock market indexes.

Q 7.11. What are the two main functions of brokerage firms?

Q 7.12. How does a prime broker differ from a retail broker?

Q 7.13. Is your rate of return higher if you short a stock in the perfect world or in the real world? Why?

Exchanges and Non-Exchanges

A retail broker would route your transaction to a centralized trading location. The most prominent exchanges are exchanges. An exchange is a centralized trading location where financial securities are traded. The two most important stock exchanges in the United States are the New York Stock Exchange (NYSE, also nicknamed the Big Board, established in 1792) and NASDAQ (originally an acronym for “National Association of Securities Dealers Automated Quotation System,” established in 1971). The NYSE used to be exclusively an auction market, in which one designated specialist (assigned for each stock) managed the auction process by trading with individual brokers on the floor of the exchange. This specialist was often a monopolist. However, even the NYSE now conducts much of its trading electronically. In contrast to the NYSE’s hybrid human-electronic process primarily in one physical location on Wall Street, NASDAQ has always been a purely electronic exchange without specialists. (For security reasons, its location—well, the location of its many computer systems—is secret.) For each NASDAQ stock, there is at least one market maker, a broker-dealer who has agreed to stand by continuously to offer to buy or sell shares—electronically of course—thereby creating a liquid and immediate market for the general public.

Shorting is like borrowing and then issuing securities. The interest on the proceeds may be earned by the broker or by the client (or be shared).

Shorting the market? Super easy!

The two big stock exchanges are the NYSE and NASDAQ. The NYSE is a hybrid market. The NASDAQ is solely electronic.
Moreover, market makers are paid for providing liquidity: They receive additional rebates from the exchange when they post a bid (short for bid price) or an ask (short for ask price) that is executed. (Outside investors can buy at least 100 shares at the quoted ask price or sell 100 shares at the quoted bid price. The ask price is always higher than the bid price.)

Most NASDAQ stocks have multiple market makers, drawn from a pool of about 500 trading firms (such as J.P. Morgan or ETrade), which compete to offer the best price. Market makers have one advantage over the general public. They can see the limit order book, which contains as-yet-unexecuted orders from investors to purchase or sell if the stock price changes—giving them a good idea at which price a lot of buying or selling activity will occur. The NYSE is the older exchange, and for historical reasons, is the most important exchange for trading most “blue chip” stocks. (“Blue chip” now means “well-established and serious.”) Ironically, the term itself came from poker, where the highest-denomination chips were blue. In 2016, the NYSE listed just under 3,000 companies worth about $20 trillion (worth about half the annual U.S. GDP). On a typical day, about $170 billion change hands. NASDAQ tends to trade smaller and high-technology firms, lists about as many firms but its companies’ market cap and dollar trading is only half that of the NYSE’s. Some stocks are traded on both exchanges.

Continuous trading—trading at any moment an investor wants to execute—relies on the presence of the standby intermediaries (specialists or market makers), who are willing to absorb shares when no one else is available. This is risky business, and thus any intermediary must earn a good rate of return to be willing to do so. To avoid this cost, some countries have organized their exchanges into noncontinuous auction systems, which match buy and sell orders a couple of times each day. The disadvantage is that you cannot execute orders immediately but have to delay until a whole range of buy and sell orders have accumulated. The advantage is that this eliminates the risk that an (expensive) intermediary would otherwise have to bear. Thus, auctions generally offer lower trading costs but slower execution.

In the United States, innovation and change in stock trading are everywhere. For example, electronic communication networks (ECNs) have made big inroads into the trading business, often replacing exchanges, especially for large institutional trades. (They can trade the same stocks that exchanges are trading, and thus they compete with exchanges in terms of cost and speed of execution.) An ECN cuts out the specialist, allowing investors to post price-contingent orders themselves. ECNs may specialize in lower execution costs, higher broker kickbacks, or faster execution. The biggest ECNs are Archipelago and Instinet. In 2005, the NYSE merged with Archipelago and converted itself from a non-profit owned by its traders into a for-profit, and NASDAQ bought Instinet. It is hard to keep track of the most recent trading arrangements. For example, in 2006, the NYSE also merged with ArcaEx, yet another electronic trading system, and merged with Euronext (a pan-European stock exchange then based in Paris) in 2007 to become NYSE Euronext. In 2012, the whole exchange was acquired by the Intercontinental Exchange (ICE), a futures broker from Atlanta. Who knows who will own what in 5 years? It may not even matter—a lot of trading has become rather opaque, happening in-house instead of in-public. Even the NYSE is not really that important and irreplaceable these days any more. It may well be eclipsed or even disappear some day.

An even more interesting venue to buy and trade stocks are crossing systems, such as ITG’s POSIT. ITG focuses primarily on matching large institutional trades with one another in an auction-like manner. If no match on the other side is found, the order may simply not be executed. But if a match is made, by cutting out the specialist or market maker, the execution is a lot cheaper than it would have been on an exchange. Recently, even more novel trading places have sprung up. For example, Liquidity uses peer-to-peer networking—like the original Napster—to match buyers and sellers in real time. ECNs and electronic limit order books are now the dominant trading systems for equities worldwide, with only the U.S. exchange floors as holdouts. Similar exchanges and computer programs are also used to trade futures, derivatives, currencies, and even some bonds.
There are many other financial markets, too. There are financial exchanges handling stock options, commodities, insurance contracts, and so on. A huge segment is the over-the-counter (OTC) markets. Over-the-counter means “call around, usually to a set of traders well known to trade in the asset, until you find someone willing to buy or sell at a price you like.” Though undergoing rapid institutional change, most bond transactions are still over-the-counter. Although OTC markets handle significantly more bond trading in terms of transaction dollar amounts than bond exchanges, OTC transaction costs are prohibitively high for retail investors. If you call without knowing the market in great detail, the person on the other end of the line will be happy to quote you a shamelessly high price, hoping that you do not know any better. The NASD (National Association of Securities Dealers) also operates a semi-OTC market for the stocks of smaller firms, which are listed on the so-called pink sheets. Foreign securities trade on their local national exchanges, but the costs for U.S. retail investors are again often too high to make direct participation worthwhile.

Q 7.14. How does a crossing system differ from an electronic exchange?

Q 7.15. What is a specialist? What is a market maker? When trading, what advantage do the two have over you?

Q 7.16. Describe some alternatives to trading on the main stock exchanges.

Investment Companies and Vehicles

In 1933/1934, Congress established the U.S. Securities and Exchange Commission (SEC) through the Securities Exchange Acts. The SEC regulates investment advisors and funds according to the Investment Advisers Act of 1940. In practice, this has allowed three different types of regulated investment companies to operate in the public markets: open-end funds, closed-end funds, and unit investment trusts (UITs).

In the United States, open-end fund is a synonym for mutual fund. (Elsewhere, mutual funds can include other classes.) Being open end means that the fund can create shares at will. Investors can also redeem their fund shares at the end of each trading day in exchange for the net asset value (NAV), which must be posted daily. This gives investors little reason to sell their fund shares to other investors—that is, mutual funds do not trade on any exchanges. The redemption right gives the law of one price. A lot of bite—fund shares are almost always worth nearly exactly what their underlying holdings are worth. If an open-end fund’s share price were to fall much below the value of its holdings, an arbitrageur could buy up the fund shares, redeem them, and thereby earn free money. (One discrepancy is due to some odd tax complications: the fund’s capital gains and losses are passed through to the fund investors at the end of every year, but they may not be what every investor experienced.) Interestingly, in the U.S. financial markets, there are now many more stock funds than individual stocks.

In a closed-end fund, there is one big initial primary offering of fund shares, and investors cannot redeem their fund shares for the underlying value. The advantage of a closed-end fund is that it can itself invest in assets that are less liquid. After all, it may not be forced to sell its holdings on the whims of its own investors. Many closed-end funds are exchange-traded, so that if a closed-end fund investor needs cash, she can resell her shares. The disadvantage of the closed-end scheme is that the law of one price has much less bite. On average, closed-end funds trade persistently below the value of their underlying holdings, roughly in line with the (often high) fees that the managers of many of these closed-end funds are charging.

There are also informal financial markets, especially OTC (over-the-counter).
Both mutual funds and closed-end fund managers are allowed to trade fund holdings quite actively—and many do so. Although some funds specialize in imitating common stock market indexes, many more try to guess the markets or try to be more “boutique.” Most funds are classified into a category based on their general trading motivation (such as “market timing,” or “growth” or “value,” or “income” or “capital appreciation”).

A unit investment trust (UIT) is sort of closed-end in its creation (usually through one big primary offering) and sort of open end in its redemption policies (usually accepting investor redemption requests on demand). Moreover, regulatory rules forbid UITs to trade actively (although this is about to change), and UITs must have a fixed termination date (even if it is 50 years in the future). UITs can be listed on a stock exchange, which makes it easy for retail investors to buy and sell them. Some early exchange-traded funds (ETFs) were structured as UITs, although this required some additional legal contortions that allowed them to create more shares on demand. This is why ETFs are nowadays usually structured as open-end funds. ETFs can compete with mutual funds in offering a myriad of portfolio baskets. Nowadays, there are more stock ETFs and mutual funds than there are stocks themselves!

Some other investment vehicles are regulated by the SEC under different rules. The most prominent may be the American Depositary Receipt (ADR). An ADR is a passive investment vehicle that usually owns the stock of only one foreign security, held in escrow at a U.S. bank (usually the Bank of New York). ADRs make it easier for U.S. retail investors to trade in foreign securities without incurring large transaction costs. ADRs are redeemable, which gives the law of one price great bite.

There are also funds that are structured so that they do not need to register with the SEC. This means that they cannot openly advertise for new investors and are limited to fewer than 100 investors. This includes most hedge funds, venture capital funds, and other private equity funds. Many offshore funds are set up to allow foreign investors to hold U.S. stocks not only without SEC regulation, but also without ever having to tread into the domain of the U.S. IRS.

Q 7.17. What should happen if the holdings of an open-end fund are worth much more than what the shares of the fund are trading for? What should happen in a closed-end fund?

Q 7.18. What is the OTC market?

Q 7.19. What are the three main types of investment companies as defined by the SEC? Which is the best deal in a perfect market?

High-Frequency Trading (HFT)

The computerization of exchanges has brought us high-frequency traders (HFTs). They generate a lot of trading volume (and even more quotes), but their net price impact for buy-and-hold investors is probably modest. There is some debate about whether HFTs add liquidity to the financial markets or siphon it off. Because their activity is both anonymous and dark to independent academic inquiry, we can only guess.

Your first question should be “How can HFTs make money in a competitive market?” The answer is that if there are non-HFT traders, such as retail investors that have posted limit orders and do not revise them quickly enough when new information arrives, then the fastest HFT can pick off their limit offers. See, markets are not really competitive in nano-seconds. There is only one HFT at the very nano-second it pounces on a standing limit order, even if another trader would have loved to pay an epsilon more a nano-second later. Thus, there has been an arms race among HFTs to be faster than others—and the speed of light has literally become the constraint! Some observers have suggested bunching orders into auctions once every second, but it is hard
7.2. Overview of Equity-Related Market Institutions

161

to know what the optimum here is (1ns? 1ms? 1s? 1h?). Besides, even if HFTs were a serious problem, it is not clear whether government intervention would improve or worsen the situation. (And, as more and more HFTs have come into the market, they have begun to compete and adding more liquidity, simply by being present.)

How Securities Appear and Disappear

Inflows

Most publicly traded equities appear on public exchanges, almost always NASDAQ, through initial public offerings (IPOs). This is an event in which a privately traded company first sells shares to ordinary retail and institutional investors. IPOs are usually executed by underwriters (investment bankers such as Goldman Sachs or Bank of America’s Merrill Lynch), which are familiar with the complex legal and regulatory process and have easy access to an investor client base to buy the newly issued shares. Shares in IPOs are typically sold at a fixed price—and for about 10% below the price at which they are likely to trade on the first day of after-market open trading. (Many IPO shares are allocated to the brokerage firm’s favorite customers, and they can be an important source of profit.)

Trading Volume in the Tech Bubble

During the tech bubble of 1999 and 2000, IPOs appreciated by 65% on their opening day on average. Getting an IPO share allocation was like getting free money. Of course, ordinary investors rarely received any such share allocations—only the underwriter’s favorite clients did. This later sparked a number of lawsuits, one of which revealed that Credit Suisse First Boston (CSFB) allocated shares of IPOs to more than 100 customers who, in return for IPO allocations, funneled between 33% and 65% of their IPO profits back to CSFB in the form of excessive trading of other stocks (like Compaq and Disney) at inflated trading commissions.

How important was this “kickback” activity? In the aggregate, in 1999 and 2000, underwriters left about $66 billion on the table for their first-day IPO buyers. If investors rebated 20% back to underwriters in the form of extra commissions, this would amount to $13 billion in excessive underwriter profits. At an average commission of 10 cents per share, this would require 130 billion shares to be traded, or an average of 250 million shares per trading day. This figure suggests that kickback portfolio churning may have accounted for as much as 10% of all shares traded!

Ritter and Welch (2002)

Usually, about a third of the company is sold in the IPO, and the typical IPO offers shares worth between $20 million and $100 million, although some are much larger (e.g., privatizations, like British Telecom). About two-thirds of all such IPO companies never amount to much or even die within a couple of years, but the remaining third soon thereafter offer more shares in seasoned equity offerings (SEOs). These days, however, much expansion in the number of shares in publicly traded companies—especially for large companies—comes not from seasoned equity offerings but from employee stock option plans, which eventually become unrestricted publicly traded shares.

The SEC is also in charge of regulating some behavior of publicly traded companies. This includes how they conduct their IPOs. It also describes how they have to behave thereafter. For example, publicly traded companies must regularly report their financials and some other information. Moreover, Congress has banned insider trading on unreleased specific information, although more general informed trading by insiders is legal (and seems to be done fairly commonly and profitably). Moreover, there are loopholes that allow smart CEOs and politicians to trade legally on inside information. (These do not apply to funds and external investors.) The SEC can only pursue civil fines. If there is fraud involved, then it is up to the states to pursue
A reverse merger has become another common way to enter the public financial markets.

Money flows out from the financial markets via dividends and share repurchases.

Shares can also shrink out of the financial markets in bankruptcies, liquidations, and delistings.

Q 7.20. What are the main mechanisms by which money flows from investors into firms?
Q 7.21. What are the institutional mechanisms by which funds disappear from the public financial markets back into the pockets of investors?
Q 7.22. How do shares disappear from the stock exchange?
Summary

This chapter covered the following major points:

- Exhibits 7.4 and 7.6 showed an analysis of historical rate of return patterns of investments in U.S. cash, bonds, stock indexes, and individual stocks.
  - Stocks, on average, had higher average rates of return than long-term bonds, which in turn had higher average rates of return than cash investments.
  - Individual stocks were riskiest. (Large-firm-type) stock market portfolios had lower risk than individual stock holdings. Bonds had modestly lower risk. Cash was least risky.
  - Stocks have outperformed cash by more than 5% per annum. However, they have only modestly outperformed long-term bonds.
- Stocks (and many other investments) have tended to correlate positively. When the stock market overall has had a good year, most individual stocks have also tended to have a good year (and vice-versa). No one knows why, but long-term bonds have tended to correlate negatively with the stock market over the last few decades (but not before).
- Most finance assumes that statistics are known. This is a leap of faith. In real life, historical data can help you in predicting the future, but it is not perfect. Historical risks and correlations are good predictors of their future equivalents; historical means may not be.
- Section 7.2 explained many institutional arrangements governing publicly traded equity securities. This includes the roles of retail and prime brokers, exchanges, and funds. It also described how stocks can be shorted, and how funds flow in and out of the financial markets.

Keywords


Answers

Q 7.1 A time-series graph shows how individual years matter. This can no longer be seen in a histogram.
Q 7.2 A histogram makes it easier to see how frequently different types of outcomes are—and thus, where the distribution is centered and how spread out it is.
Q 7.3 A compound return graph shows how a time series of rates of return interacts to produce long-run returns. In other words, you can see whether a long-run investment would have made or lost money. This is difficult to see in a time-series graph.
Q 7.4 Note that because the returns in (b) and (c) alternate, you just need to work out the safe two-year returns—thereafter, they will continue in their (unrealistic) patterns.

1. 5% for both.
2. Over two years, you earn $1.00 \cdot 1.10 - 1 = 10.00\%$. This means
that the annualized rate of return is \( \sqrt{1.10} - 1 \approx 4.88\% \). This is lower than the average rate of return, which is still 5%.

3. Over two years, you earn 0.9 \cdot 1.20 - 1 = 8.00\%. This means that the annualized rate of return is \( \sqrt{1.08} - 1 \approx 3.92\% \). This is lower than the 5% average rate of return.

Yes. The difference between its annualized and its average rate of return is greater for a more volatile investment.

Q 7.5 The difference is 0, because the risk-free rate has no standard deviation.

Q 7.6 The risk is usually increasing: lowest for cash, then bonds, then the stock market portfolio, and finally individual stocks. The average reward is increasing for the first three, but this is not necessarily true for an individual stock.

Q 7.7 Usually (but not always), individual stocks are riskier.

Q 7.8 Yes. For example, look at UAL in Exhibit 7.6. It lost everything but still had a positive average arithmetic rate of return.

Q 7.9 To graph the market beta, the rate of return on the market (e.g., the S&P 500) should be on the x-axis, and the rate of return on the investment for which you want to determine the market beta should be on the y-axis. A data point is the two rates of return from the same given time period (e.g., over a year). The market beta is the slope of the best-fitting line.

Q 7.10 The market beta of the market is 1—you are plotting the rate of return on the market on both the x-axis and the y-axis, so the beta is the slope of this 45° diagonal line.

Q 7.11 Brokers execute orders and keep track of investors’ portfolios. They also facilitate purchasing on margin.

Q 7.12 Prime brokers are usually used by larger investors. Prime brokers allow investors to employ their own traders to execute trades. (Like retail brokers, prime brokers provide portfolio accounting, margin, and securities borrowing.)

Q 7.13 Your rate of return is higher if you short a stock in the perfect world because you earn interest on the proceeds. In the real world, your broker may help himself to this interest.

Q 7.14 A crossing system does not execute trades unless there is a counterparty. It also tries to cross orders a few times a day.

Q 7.15 The specialist is often a monopolist who makes the market on the NYSE. The specialist buys and sells from her own inventory of a stock, thereby “making a market.” Market makers are the equivalent on NASDAQ, but there are usually many and they compete with one another. Unlike ordinary investors, both specialists and market makers can see the limit orders placed by other investors.

Q 7.16 The alternatives are often electronic, and they often rely on matching trades—thus, they may not execute trades that they cannot match. Electronic communication networks are the dominant example of these. Another alternative is to execute the trade in the over-the-counter (OTC) market, which is a network of geographically dispersed dealers who are making markets in various securities.

Q 7.17 In an open-ended fund, you should buy fund shares and request redemption. (You could short the underlying holdings during the time you wait for the redemption in order not to suffer price risk.) In a closed-ended fund, you would have to oust the management to allow you to redeem your shares.

Q 7.18 The OTC is not really a market. Instead, it simply means that traders handle transactions on a one-on-one basis.

Q 7.19 UITs, open-ended funds (mutual funds), and closed-ended investment funds. In a perfect market, none is the best deal. You always get what you pay for.

Q 7.20 The main mechanisms by which money flows from investors into firms are (a) IPOs and SEOs, and (b) reverse mergers, which are then sold off to investors.

Q 7.21 Funds disappear from the public financial markets back into the pockets of investors through dividends and share repurchases.

Q 7.22 Shares can disappear in a delisting or a repurchase.
End of Chapter Problems

Q 7.23. Using the information in Exhibit 7.4, compute the discrepancy between arithmetic and geometric rates of return for cash and stocks. Which one is lower? Why?

Q 7.24. Broadly speaking, what was the average rate of return on cash, bonds, and stocks? What time period are your numbers from?

Q 7.25. Broadly speaking, what was the average risk of cash, bonds, and stocks? What time period are your numbers from?

Q 7.26. How good are historical statistics as indicators of future statistics? Which kinds of statistics are better? Which kinds are worse?

Q 7.27. Does the market beta of stocks in the market average out to zero?

Q 7.28. Give an example in which a stock had a positive average rate of return, even though it lost its investors' money.

Q 7.29. Are stock funds or bond funds that quote historical average rates of return more misleading? Would you have ended up with more money in a stock fund or a bond fund if they both quoted similar historical mean rate of return performances?

Q 7.30. Looking at the figures in this chapter, did 20-year bonds move with or against the U.S. stock market?

Q 7.31. Do individual stocks tend to move together? How could this be measured?

Q 7.32. Explain the differences between a market order and a limit order.

Q 7.33. What extra function do retail brokers handle that prime brokers do not?

Q 7.34. Describe the differences between the NYSE and NASDAQ.

Q 7.35. Roughly, how many firms are listed on the NYSE? How many are listed on NASDAQ? Then use a financial website to find an estimate of the current number.

Q 7.36. Is NASDAQ a crossing market?

Q 7.37. What are the two main mechanisms by which a privately held company can go public?

Q 7.38. When and under what circumstances was the SEC founded?

Q 7.39. Insider trading is a criminal offense. Does the SEC prosecute these charges to put violators behind bars?

Q 7.40. What is the OTC market?

Q 7.41. If a firm repurchases 1% of its shares, does this change the capitalization of the stock market on which it lists? If a firm pays 1% of its value in dividends, does this change the capitalization of the stock market on which it lists?

---

Data and Programming for Masters Students

Please see Page 36 for a description of stock price data vendors.

**Task A:** Repeat Exhibit 7.4 for (a) a portfolio of “value” firms; (b) “growth” firms; (c) “high-dividend” firms; (d) “Fortune-500” firms (in time); (e) Gold and/or Precious Metals; (f) German stocks (in dollars); (g) British stocks; (h) Japanese stocks; (i) Deutsche Bundesanleihen (i.e., German government bonds); and (j) Real-estate. Because data about the rates of return on these assets are not easily available, use (preferably open-ended) mutual fund (or ETF) return data instead.

**Task B:** Calculate your own equal-weighted and value-weighted stock market index returns from the constituent stock returns over the last 5 years. (For the value-weighted index, you must use ex-ante marketcap data. For kicks, you can try out what happens when you use ex-post marketcap data.) Compare your indexes, the Dow-Jones 30 index, the S&P 500 index, and the CRSP equal- and value-weighted indexes. Does it matter with what frequency you are updating or not updating the lagged marketcap?