Factor Theory

Andrew Ang

DOI:10.1093/acprof:oso/9780199959327.003.0006

Abstract and Keywords

Assets earn risk premiums because they are exposed to underlying factor risks. The capital asset pricing model (CAPM), the first theory of factor risk, states that assets that crash when the market loses money are risky and therefore must reward their holders with high risk premiums. While the CAPM defines bad times as times of low market returns, multifactor models capture multiple definitions of bad times across many factors and states of nature.

Keywords: CAPM, multifactor model, APT, factor risk premium, diversification, efficient frontier, capital market line, mean-variance efficient, tangency portfolio, pricing kernel, stochastic discount factor, near-efficient markets, EMH

Chapter Summary

Assets earn risk premiums because they are exposed to underlying factor risks. The capital asset pricing model (CAPM), the first theory of factor risk, states that assets that crash when the market loses money are risky and therefore must reward their holders with high risk premiums. While the CAPM defines bad times as times of low market returns,
multifactor models capture multiple definitions of bad times across many factors and states of nature.

1. The 2008–2009 Financial Crisis
During the financial crisis of 2008 and 2009, the price of most risky assets plunged. Table 6.1 shows that U.S. large cap equities returned -37%; international and emerging markets equities had even larger losses. The riskier fixed income securities, like corporate bonds, emerging market bonds, and high yield bonds, also fell, tumbling along with real estate. “Alternative” investments like hedge funds, which trumpeted their immunity to market disruptions, were no safe refuge: equity hedge funds and their fixed income counterparts fell approximately 20%. Commodities had losses exceeding 30%. The only assets to go up during 2008 were cash (U.S. Treasury bills) and safe-haven sovereign bonds, especially long-term U.S. Treasuries.

Table 6.1 Returns of Asset Classes in 2008

<table>
<thead>
<tr>
<th>Asset Type</th>
<th>Index/Portfolio</th>
<th>Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash</td>
<td>Three-month T-bill</td>
<td>1.3%</td>
</tr>
<tr>
<td>Core Bonds</td>
<td>Barcap Aggregate Index</td>
<td>5.2%</td>
</tr>
<tr>
<td>Global Bonds</td>
<td>Citigroup World Government</td>
<td>10.9%</td>
</tr>
<tr>
<td>TIPS</td>
<td>Citigroup US Inflation Linked</td>
<td>-1.2%</td>
</tr>
<tr>
<td>Emerging Market Bonds</td>
<td>JPM Emerging Markets Bond Index</td>
<td>-9.7%</td>
</tr>
<tr>
<td>US High Yield</td>
<td>Merrill Lynch High Yield Master</td>
<td>-26.3%</td>
</tr>
<tr>
<td>Large Cap Equity</td>
<td>S&amp;P 500</td>
<td>-37.0%</td>
</tr>
<tr>
<td>Small Cap Equity</td>
<td>Russell 2000</td>
<td>-33.8%</td>
</tr>
<tr>
<td>International Equity</td>
<td>MSCI World ex US</td>
<td>-43.2%</td>
</tr>
<tr>
<td>Emerging Markets Equity</td>
<td>IFC Emerging Markets</td>
<td>-53.2%</td>
</tr>
<tr>
<td>Public Real Estate</td>
<td>NAREIT Equity REITS</td>
<td>-37.7%</td>
</tr>
<tr>
<td>Private Real Estate</td>
<td>NCREIF Property Index</td>
<td>-16.9%</td>
</tr>
<tr>
<td>Private Capital</td>
<td>Venture Economics (Venture and Buyouts)</td>
<td>-20.0%</td>
</tr>
</tbody>
</table>
Why did so many asset classes crash all at once? And given that they did, was the concept of diversification dead?

In this chapter, we develop a theory of factor risk premiums. The factor risks constitute different flavors of bad times and the investors who bear these factor risks need to be compensated in equilibrium by earning factor risk premiums. Assets have risk premiums not because the assets themselves earn risk premiums; assets are bundles of factor risks, and it is the exposures to the underlying factor risks that earn risk premiums. These factor risks manifest during bad times such as the financial crisis in late 2008 and early 2009.

Factors are to assets what nutrients are to food. Table 6.2 is from the Food and Nutrition Board, which is part of the Institute of Medicine of the National Academies, and lists recommended intakes of the five macronutrients—water, carbohydrates, protein, fiber, and fat—for an “average” male, female, and child. Carbohydrates can be obtained from food made from cereals and grains. Protein is obtained from meat and dairy products. Fiber is available from wheat and rice. Fat we can consume from animals but also certain plant foods such as peanuts. Each type of food is a bundle of nutrients. Many foods contain more than just one macronutrient: for example, rice contains both carbohydrates and fiber. Different individuals, whether sick or healthy, male or female, or young or old, have different macronutrient requirements. We eat food for the underlying nutrients; it is the nutrients that give sustenance.
<table>
<thead>
<tr>
<th>Macronutrients</th>
<th>Male</th>
<th>Female</th>
<th>Child</th>
<th>Examples of Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>3.7L/day</td>
<td>2.7L/day</td>
<td>1.7L/day</td>
<td>Bread, Beans, Potato Rice</td>
</tr>
<tr>
<td>Carbohydrates</td>
<td>130g/day</td>
<td>130g/day</td>
<td>130g/day</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>56g/day</td>
<td>46g/day</td>
<td>19g/day</td>
<td>Cheese, Milk, Fish, Soya bean</td>
</tr>
<tr>
<td>Fiber</td>
<td>38g/day</td>
<td>25g/day</td>
<td>25g/day</td>
<td>Peas, Wheat, Rice</td>
</tr>
<tr>
<td>Fat</td>
<td>20 – 35% of calories</td>
<td>25 – 35% of calories</td>
<td></td>
<td>Oily fish, Peanuts, Animal fat</td>
</tr>
</tbody>
</table>

Factor risks are the driving force behind assets’ risk premiums. An important theory of factor risk is the CAPM, which we explore in the next section. The CAPM (p.195) states that there is only one factor driving all asset returns, which is the market return in excess of T-bills. All assets have different exposures to the market factor and the greater the exposure, the higher the risk premium. The market is an example of a tradeable, investment factor. Other examples include interest rates, value-growth investing, low volatility investing, and momentum portfolios. Factors can also be fundamental macro factors, like inflation and economic growth. Assets have different payoffs during high or low inflation periods or during economic recessions and expansions. We leave a complete exposition of the various types of factors to the next chapter. In this chapter, we describe the underlying theory of factor risk.

There are three similarities between food and assets:

1. Factors matter, not assets.
   If an individual could obtain boring, tasteless nutrients made in a laboratory, she would comfortably meet her nutrient requirements and lead a healthy life. (She would, however, deprive herself of gastronomic enjoyment.) The factors behind the assets matter, not the assets themselves. Investing right requires looking through asset class labels to understand the factor content, just as eating right requires looking through food labels to understand the nutrient content.

2. Assets are bundles of factors.
   Foods contain various combinations of nutrients. Certain foods are nutrients themselves—like water—or are close to containing only one type of nutrient, as in the case of rice for carbohydrates. But generally foods contain many nutrients. Similarly, some asset classes can be considered factors themselves—like equities and government fixed income securities—while other assets contain many different factors. Corporate bonds, hedge funds, and private equity contain different amounts of equity risk, volatility risk, interest rate risk, and default risk. Factor theory predicts these assets have risk premiums that reflect their underlying factor risks.
3. Different investors need different risk factors. Just as different people have different nutrient needs, different investors have different optimal exposures to different sets of risk factors. Volatility, as we shall see, is an important factor. Many assets and strategies lose money during times of high volatility, such as observed during the 2007–2008 financial crisis. Most investors dislike these times and would prefer to be protected against large increases in volatility. A few brave investors can afford to take the opposite position; these investors can weather losses during bad times to collect a volatility premium during normal times. They are paid risk premiums as compensation for taking losses—sometimes big losses, as in 2008–2009—during volatile times.

Another example is that investors have different desired exposure to economic growth. One investor may not like times of shrinking GDP growth because he is likely to become unemployed in such circumstances. Another investor—a bankruptcy lawyer, perhaps—can tolerate low GDP growth because his labor income increases during recessions. The point is that each investor has different preferences, or risk aversion coefficients, for each different source of factor risk. There is one difference, however, between factors and nutrients. Nutrients are inherently good for you. Factor risks are bad. It is by enduring these bad experiences that we are rewarded with risk premiums. Each different factor defines a different set of bad times. They can be bad economic times—like periods of high inflation and low economic growth. They can be bad times for investments—periods when the aggregate market or certain investment strategies perform badly. Investors exposed to losses during bad times are compensated by risk premiums in good times. The factor theory of investing specifies different types of underlying factor risk, where each different factor represents a different set of bad times or experiences. We describe the theory of factor risk by starting with the most basic factor risk premium theory—the CAPM, which specifies just one factor: the market portfolio.

3. CAPM
The CAPM was revolutionary because it was the first cogent theory to recognize that the risk of an asset was not how that asset behaved in isolation but how that asset moved in relation to other assets and to the market as a whole. Before the CAPM, risk was often thought to be an asset’s own volatility. The CAPM said (p.197) this was irrelevant and that the relevant measure of risk was how the asset covaried with the market portfolio—the beta of the asset. It turns out that asset volatility itself matters, as we shall see in chapter 7, but for the purpose of describing the CAPM and its incredible implications, we can ignore this for the time being.

The CAPM was formulated in the 1960s by Jack Treynor (1961), William Sharpe (1964), John Lintner (1965), and Jan Mossin (1966), building on the principle of diversification and mean-variance utility introduced by Harry Markowitz in 1952. For their work on CAPM and portfolio choice, Sharpe and Markowitz received the 1990 Nobel Prize in economics. (Merton Miller was awarded the Nobel Prize the same year for contributions to corporate finance.) Lintner and Mossin, unfortunately, had both died by then. Treynor, whose original manuscript was never published, has never received the recognition that he deserved.

I state upfront that the CAPM is well known to be a spectacular failure. It predicts that asset risk premiums depend only on the asset’s beta and there is only one factor that matters, the market portfolio. Both of these predictions have been demolished in thousands of empirical studies. But, the failure has been glorious, opening new vistas of understanding for asset owners who must hunt for risk premiums and manage risk.

The basic intuition of the CAPM still holds true: that the factors underlying the assets determine asset risk premiums and that these risk premiums are compensation for investors’ losses during bad times. Risk is a property not of an asset in isolation but how the assets move in relation to each other. Even though the CAPM is firmly rejected by data, it remains the workhorse model of finance: 75% of finance professors advocate using it, and 75% of CFOs employ it in actual capital budgeting decisions despite the fact that the CAPM does not hold.\(^1\) It works approximately, and well enough for most applications, but it fails miserably in certain situations (as the
next chapter will detail). Part of the tenacious hold of the CAPM is the way that it conveys intuition of how risk is rewarded.

What does the CAPM get right?

3.1. CAPM Lesson 1: Don’t Hold an Individual Asset, Hold the Factor

The CAPM states that one factor exists and that factor is the market portfolio, where each stock is held in proportion to its market capitalization. This corresponds to a market index fund. The factor can be optimally constructed by holding many assets so that nonfactor, or idiosyncratic risk, is diversified away. Asset owners are better off holding the factor—the market portfolio—than individual stocks. Individual stocks are exposed to the market factor, which carries the risk premium (it is the nutrient), but also have idiosyncratic risk, which is not rewarded by a risk premium (this is the part that carries no nutritional value). Investors can diversify away the idiosyncratic part and increase their returns by holding the market factor portfolio, rather than any other combination of individual stocks. The market portfolio represents systematic risk, and it is pervasive: all risky assets have risk premiums determined only by their exposure to the market portfolio. Market risk also affects all investors, except those who are infinitely risk averse and hold only risk-free assets.

The key to this result is diversification. The CAPM is based on investors having mean-variance utility and, as chapter 3 shows, the most important concept in mean-variance investing is diversification. Diversification ensures that, absent perfect correlation, when one asset performs badly, some other assets will perform well, and so gains can partly offset losses.

Investors never want to hold assets in isolation; they improve their risk-return trade-off by diversifying and holding portfolios of many assets. This balance across many assets that are not perfectly correlated improves Sharpe ratios.

Investors will diversify more and more until they hold the most diversified portfolio possible—the market portfolio. The market factor is the best, most-well diversified portfolio investors can hold under the CAPM.
The CAPM states that the market portfolio is held by every investor—a strong implication that is outright rejected in data. Nevertheless, it is useful to understand how we can leap from a diversified portfolio to the market being the only relevant factor.

Recall the mean-variance frontier with the capital allocation line (CAL) from chapter 3 (see Figure 3.10), which is reproduced in Figure 6.3. This is the solution to the mean-variance investing problem. Investors hold different amounts of the risk-free asset and the mean-variance efficient (MVE) portfolio depending on their risk aversion. Now here come the strong assumptions of the CAPM. Assume that the set of means, volatilities, and correlations are the same for all investors. (p.199) Then all investors hold the same MVE portfolio—just in different quantities depending on their own risk aversion. Since everyone holds the same MVE and this is the best portfolio that can be held by all investors, the MVE portfolio becomes the market factor in equilibrium.
**Equilibrium**

The equilibrium concept is extremely important. Equilibrium occurs when investor demand for assets is exactly equal to supply. The market is the factor in equilibrium because in CAPM land, everyone holds the MVE portfolio (except for those who are infinitely risk averse). If everyone’s optimal risky portfolio (which is the MVE) assigns zero weight to a certain asset, say AA stock, then this cannot be an equilibrium. Someone must hold AA so that supply equals demand. If no one wants to hold AA, then AA must be overpriced and the expected return of AA is too low. The price of AA falls. The expected payoff of AA stays constant under CAPM assumptions, so that as the price of AA falls, the expected return of AA increases. AA’s price falls until investors want to hold exactly the number of AA shares outstanding. Then, the expected return is such that supply is equal to demand in equilibrium. Since all investors hold the MVE portfolio, the MVE portfolio becomes the market portfolio, and the market consists of each asset in terms of market capitalization weights.

Equilibrium ensures that the factor—the market portfolio—will have a risk premium and that this risk premium will not disappear. The market factor is *systematic* and affects all assets. The market risk premium is a function of the underlying investors’ risk aversions and utilities. That is, the risk premium of the market factor reflects the full setup of all people in the economy. The factors that we introduce later—tradeable factors like value-growth investing and volatility investing or macro factors like inflation and economic growth—will also carry risk premiums based on investor characteristics, the asset universe, and the production capabilities of the economy. They will disappear only if the economy totally changes. Equilibrium factor risk premiums will not disappear because clever hedge funds go and trade them—these types of investment strategies are not factors. Investors cannot arbitrage away the market factor and all other systematic factors.
3.2. CAPM Lesson 2: Each Investor Has His Own Optimal Exposure of Factor Risk

In Figure 6.3, all investors will hold the market portfolio, just in different proportions. Pictorially, they have different proportions of the risk-free asset and the market portfolio and lie on different positions on the CAL line (see chapters 2 and 3). Thus, each individual investor has a different amount of factor exposure just as different individuals have different nutrient requirements.

(p.200) 3.3. CAPM Lesson 3: The Average Investor Holds the Market

The market portfolio represents the average holdings across investors. The intersection of the CAL with the mean-variance frontier represents an investor who holds 100% in the MVE portfolio. This tangency point represents the average investor. The risk aversion corresponding to this 100% portfolio position is the risk aversion of the market.3

Note that as investors differ from the average investor, they will be exposed to more or less market factor risk depending on their own risk preferences. We will come back to this extensively in chapter 14 when we describe factor investing.

3.4. CAPM Lesson 4: The Factor Risk Premium Has an Economic Story

The CAL in Figure 6.3 for a single investor is called the capital market line (CML) in equilibrium, since under the strong assumptions of the CAPM every investor has the same CML. (The MVE portfolio is the market factor portfolio.) The equation for the CML pins down the risk premium of the market:

\[(6.1) \quad E(r_m) - r_f = \bar{\gamma} \sigma_m\]

where \(E(r_m) - r_f\) is the market risk premium, or the expected return on the market in excess of the risk-free rate; \(\bar{\gamma}\) is the risk aversion of the "average" investor; and \(\sigma_m\) is the volatility of the market portfolio. The CAPM derives the risk premium in terms of underlying agent preferences (\(\bar{\gamma}\) is the average risk aversion across all investors, where the average is taken weighting each individual’s degree of risk aversion in proportion to the wealth of that agent).
According to the CAPM in equation (6.1), as the market becomes more volatile, the expected return of the market increases and equity prices contemporaneously fall, all else equal. We experienced this in 2008 and 2009 when volatility skyrocketed and equity prices nosedived. Expected returns in this period on were very high (and realized returns were indeed high in 2009 and 2010). It is intuitive that the market risk premium in equation (6.1) is proportional to market variance because under the CAPM investors have mean-variance preferences: they dislike variances and like expected returns. The market portfolio is the portfolio that has the lowest volatility among all portfolios that share the same mean as the market, or the market has the highest reward-to-risk ratio (or Sharpe ratio). The market removes all idiosyncratic risk. This remaining risk has to be rewarded, and equation (6.1) states a precise equation for the risk premium of the market.

(p.201) As the average investor becomes more risk averse to variance (so \( \gamma \) increases), the risk premium of the market also increases.

3.5. CAPM Lesson 5: Risk Is Factor Exposure

The risk of an individual asset is measured in terms of the factor exposure of that asset. If a factor has a positive risk premium, then the higher the exposure to that factor, the higher the expected return of that asset.

The second pricing relationship from the CAPM is the traditional beta pricing relationship, which is formally called the security market line (SML). Denoting stock \( i \)'s return as \( r_i \) and the risk-free return as \( r_f \), the SML states that any stock’s risk premium is proportional to the market risk premium:

\[
E(r_i) - r_f = \frac{\text{cov}(r_i, r_m)}{\text{var}(r_m)} (E(r_m) - r_f) = \beta_i (E(r_m) - r_f).
\]

The risk premium on an individual stock, \( E(r_i) - r_f \), is a function of that stock’s beta, \( \beta_i = \frac{\text{cov}(r_i, r_m)}{\text{var}(r_m)} \), the beta is a measure of how that stock co-moves with the market portfolio, and the higher the co-movement (the higher \( \text{cov}(r_i, r_m) \)), the higher the asset’s beta.

I will not formally derive equation (6.2). But it contains some nice intuition: mean-variance investing is all about diversification benefits. Beta—the CAPM’s measure of risk—is
a measure of the lack of diversification potential. Beta can be written as $\beta_i = \rho_{im} \sigma_i / \sigma_m$, where $\rho_{im}$ is the correlation between asset $i$’s return and the market return, $\sigma_i$ is the volatility of the return of asset $i$, and $\sigma_m$ is the volatility of the market factor.

Chapter 3 emphasized that the lower the correlation with a portfolio, the greater the diversification benefit with respect to that portfolio because the asset was more likely to have high returns when the portfolio did badly. Thus, high betas mean low diversification benefits.

If we start from a diversified portfolio, investors find assets with high betas—those assets that tend to go up when the market goes up, and vice versa—to be unattractive. These high beta assets act like the diversified portfolio the investor already holds, and so they require high expected returns to be held by investors. In contrast, assets that pay off when the market tanks are valuable. These assets have low betas. Low beta assets have tremendous diversification benefits and are very attractive to hold. Investors, therefore, do not need to be compensated very much for holding them. In fact, if the payoffs of these low beta assets are high enough when the market return is low, investors are willing to pay to hold these assets rather than be paid. That is, assets with low enough betas actually have negative expected returns. These assets are so attractive because they have large payoffs when the market is crashing. Gold, or sometimes government bonds, are often presented as examples of low (or negative) beta assets which tend to pay off when the stock market crashes.

(Government bonds were one of the few asset classes to do well in the financial crisis in 2008. See also chapter 9 on fixed income.)

3.6. CAPM Lesson 6: Assets Paying Off in Bad Times Have Low Risk Premiums

Another way to view the SML relationship in equation (6.2) is that the risk premium in the CAPM is a reward for how an asset pays off in bad times. Bad times are defined in terms of the factor, which is the market portfolio, so bad times correspond to low (or negative) market returns. If the asset has losses when the market has losses, the asset has a high beta. When the market has gains, the high beta asset also gains in value. Investors are, on average, risk averse so that the gains during good times do not cancel out the losses.
during bad times. Thus, high beta assets are risky and require high expected returns to be held in equilibrium by investors.

Conversely, if the asset pays off when the market has losses, the asset has a low beta. This asset is attractive and the expected return on the asset can be low—investors do not need much compensation to hold these attractive assets in equilibrium. More generally, if the payoff of an asset tends to be high in bad times, this is a valuable asset to hold and its risk premium is low. If the payoff of an asset tends to be low in bad times, this is a risky asset and its risk premium must be high.

In the CAPM, the bad returns are defined as low returns of the market portfolio. This is, of course, very restrictive: there are many more factors than just the market. In multifactor models, all the intuitions of CAPM Lessons 1 through 6 hold. Except that, with multiple factors, each factor defines its own set of bad times.

4. Multifactor Models
Multifactor models recognize that bad times can be defined more broadly than just bad returns on the market portfolio. The first multifactor model was the arbitrage pricing theory (APT), developed by Stephen Ross (1976). It uses the word “arbitrage” because the factors cannot be arbitraged or diversified away—just like the single market factor in the CAPM. In equilibrium, investors must be compensated for bearing these multiple sources of factor risk. While the CAPM captures the notion of bad times solely by means of low returns of the market portfolio, each factor in a multifactor model provides its own definition of bad times.

(p.203) 4.1. Pricing Kernels
To capture the composite bad times over multiple factors, the new asset pricing approach uses the notion of a pricing kernel. This is also called a stochastic discount factor (SDF). We denote the SDF as $m$. The SDF is an index of bad times, and the bad times are indexed by many different factors and different states of nature. Since all the associated recent asset pricing theory uses this concept and terminology, it is worth spending a little time to see how this SDF approach is related to the traditional CAPM approach. There is some nice intuition that comes about from using the SDF, too. (For the less
technically inclined, you are welcome to skip the next two subsections and start again at section 4.3.)

By capturing all bad times by a single variable $m$, we have an extremely powerful notation to capture multiple definitions of bad times with multiple variables. The CAPM is actually a special case where $m$ is linear in the market return.$^5$

\begin{equation}
(6.3) \quad m = a + b \times r_m
\end{equation}

for some constants $a$ and $b$. (A pricing kernel that is linear in the market gives rise to a SML that with asset betas with respect to the market in equation (6.2).) With our “$m$” notation, we can specify multiple factors very easily by having the SDF depend on a vector of factors, $F = (f_1, f_2, ..., f_K)$:

\begin{equation}
(6.4) \quad m = a + b_1 f_1 + b_2 f_2 + ... + b_K f_K
\end{equation}

where each of the $K$ factors themselves define different bad times.

Another advantage of using the pricing kernel $m$ is that while the CAPM restricts $m$ to be linear, the world is nonlinear. We want to build models that capture this nonlinearity.$^6$ Researchers have developed some complicated forms for $m$, and some of the workhorse models that we discuss in chapters 8 and 9 describing equities and fixed income are nonlinear.
4.2. Pricing Kernels versus Discount Rate Models

Here’s how these pricing kernels work. In the traditional discount rate model of the CAPM, we find the price of asset $i$ by discounting its payoff next period back to today:

\[(6.5)\]

$$P_i = E\left[ \frac{\text{payoff}_i}{1 + E(r)} \right],$$

where the discount rate is given by $E(r) = r_f + \beta_i (E(r_m) - r_f)$ according to the CAPM. Under the SDF model, we can equivalently write the price of the asset using $m$-notation: 7

\[(6.6)\]

$$P_i = E\left[ m \times \text{payoff}_i \right],$$

and hence the name “stochastic discount factor,” because we discount the payoffs using $m$ in equation (6.6), just as we discount the payoff by the discount rate in the more traditional discount formula (6.5). The SDF is called a pricing kernel, borrowing the use of the word “kernel” from statistics, because one can estimate $m$ in equation (6.6) using a kernel estimator. Since it is a kernel that prices all assets, it is a “pricing kernel.” Students of probability and statistics will recognize that the price in equation (6.6) is an expectation taken with respect to the pricing kernel, so this gives rise to the SDF also being called the state price density.

We can divide both the right- and left-hand sides of equation (6.6) by the asset’s current price, $P_i$, to obtain

\[(6.7)\]

$$\frac{P_i}{P} = E\left[ \frac{m \times \text{payoff}_i}{P_i} \right] = E\left[ m \times (1 + r) \right].$$

A special case of equation (6.7) occurs when the payoffs are constant. That would give us a risk-free asset, so the price of a risk-free bond is simply $1/(1 + r) = E[m \times 1]$.

It turns out that we can write the risk premium of an asset in a relation very similar to the SML of the CAPM in equation (6.2): 8

\[(6.8)\]

$$E(r_i) - r_f = \frac{\text{cov}(r_i, m)}{\text{var}(m)} \left( \frac{\text{var}(m)}{E(m)} \right)$$

\[(p.205)\]

where $\beta_{im} = \text{cov}(r_i, m)/\text{var}(m)$ is the beta of the asset with respect to the SDF. Equation (6.8) captures the “bad times” intuition that we had earlier from the CAPM. Remember that $m$ is an index of bad times. The higher the
payoff of the asset is in bad times (so the higher $\text{cov}(r_p, m)$ and the higher $\beta_{wa}$), the lower the expected return of that asset. The higher beta in equation (6.8) is multiplied by the price of “bad times” risk, $\lambda_m = -\text{var}(m)/E(m)$, which is the inverse of factor risk, which is why there is a negative sign. Equation (6.8) states directly the intuition of Lesson 6 from the CAPM: higher covariances with bad times lead to lower risk premiums.

Assets that pay off in bad times are valuable to hold, so prices for these assets are high and expected returns are low.

Just as the CAPM gives rise to assets having betas with respect to the market, multiple factors in the SDF in equation (6.4) gives rise to a multi-beta relation for an asset’s risk premium:

\begin{equation}
E(r) = r_f + \beta_{1f}E(f_1) + \beta_{2f}E(f_2) + \ldots + \beta_{kf}E(f_k),
\end{equation}

where $\beta_{ik}$ is the beta of asset $i$ with respect to factor $k$ and $E(f_k)$ is the risk premium of factor $k$. For macro factors, $f_1$ could be inflation and $f_2$ could be economic growth, for example. Bad times are characterized by times of high inflation, low economic growth, or both. For an example for multiple investment factors, $f_1$ could be the market portfolio and $f_2$ could be an investing strategy based on going long value stocks and short growth stocks. Value stocks outperform growth stocks in the long run (see chapter 7). Bad times are characterized by low market returns, value stocks underperforming growth stocks, or both.

### 4.3. Multifactor Model Lessons

The key lessons in the multifactor world are in fact the same from the CAPM:

<table>
<thead>
<tr>
<th>CAPM (Market Factor)</th>
<th>Multifactor Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1 Diversification works.</td>
<td>Diversification works.</td>
</tr>
<tr>
<td>The market diversifies away idiosyncratic risk.</td>
<td>The tradeable version of a factor diversifies away idiosyncratic risk.</td>
</tr>
<tr>
<td>Lesson 2 Each investor has her own optimal exposure of the market portfolio.</td>
<td>Each investor has her own optimal exposure of each factor risk.</td>
</tr>
</tbody>
</table>
### CAPM (Market Factor) vs. Multifactor Models

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>The average investor holds the market.</td>
<td>The average investor holds the market.</td>
</tr>
<tr>
<td>4</td>
<td>The market factor is priced in equilibrium under the CAPM assumptions.</td>
<td>Risk premiums exist for each factor assuming no arbitrage or equilibrium.</td>
</tr>
<tr>
<td>5</td>
<td>Risk of an asset is measured by the CAPM beta.</td>
<td>Risk of an asset is measured in terms of the factor exposures (factor betas) of that asset.</td>
</tr>
<tr>
<td>6</td>
<td>Assets paying off in bad times when the market return is low are attractive, and these assets have low risk premiums.</td>
<td>Assets paying off in bad times are attractive, and these assets have low risk premiums.</td>
</tr>
</tbody>
</table>

(p.206) The $64,000 question with multifactor pricing kernel models is: how do you define bad times? For the average investor who holds the market portfolio, the answer is when an extra $1 becomes very valuable. This interprets the SDF as the marginal utility of a representative agent. We will come back to this formulation in chapter 8 when we characterize the equity risk premium. Times of high marginal utility are, for example, periods when you’ve just lost your job so your income is low and any extra dollars are precious to you. Your consumption is also low during these times. In terms of the average, representative consumer, this also corresponds to a macro factor definition of a bad time: bad times are when GDP growth is low, consumption is low, or economic growth in general is low. Times of high marginal utility could also be defined in relative terms: it could be when your consumption is low relative to your neighbor or when your consumption is low relative to your past consumption. In chapter 2, we captured the former using a catching up with the Joneses utility function and the latter with a habit utility function.

During 2008–2009, the financial crisis was a bad time with high volatility and large financial shocks. So volatility is an important factor, and the next chapter shows that many risky assets perform badly when volatility is high. Factors can also be tradeable, investment styles. Some of these include liquid,
public market asset classes like bonds and listed equities. Others include investment styles that are easily replicable and can be implemented cheaply (but often are not when they are delivered to customers) and in scale, like value/growth strategies.\(^9\)

**Failure of the CAPM**

The CAPM is derived using some very strong assumptions. It’s worth taking a moment to examine these assumptions and discuss what happens when they are relaxed.

1. Investors have only financial wealth. As Part I of this book has emphasized, investors have unique income streams and liabilities, and their optimal portfolio choice has to take these into consideration. Liabilities are often denominated in real terms—we want to maintain a standard of living even if prices rise, for example. Income streams are usually risky, and income declines during periods of low economic growth. This makes variables like inflation and growth important factors because many investors’ income and liabilities change as the macro variables change. One particular important factor that drives asset returns is human capital, or labor income risk.\(^10\) In an influential paper, Jagannathan and Wang (1996) found large improvements in the performance of the CAPM when labor income risk is taken into account.

2. Investors have mean-variance utility. As chapter 2 emphasizes, more realistic utility functions often have an asymmetric treatment of risk because investors are generally more distressed by losses than pleased by gains. We should expect, then, to find deviations from the CAPM among stocks that have different measures of downside risk. Ang, Chen, and Xing (2006) show that stocks with greater downside risk have higher returns. A large number of papers show that other higher moment risk, like skewness and kurtosis, also carry risk premiums.\(^11\)

3. Single-period investment horizon. By itself an investment horizon of one period is a minor assumption. Merton (1971, 1973) provides a famous extension of the CAPM to the dynamic case. In this setting, the CAPM holds in each single period.
While the long investment horizon is an inconsequential assumption for the CAPM theory, there is a huge implication when we extend portfolio choice to a dynamic, long-horizon setting. As Part I of this book has shown, the optimal strategy for long-term investors is to rebalance (see chapter 3).  

4. Investors have *homogeneous* expectations. 

This assumption ensures that all investors hold the same MVE portfolio in the CAPM world and that, in equilibrium, the MVE portfolio becomes the market portfolio. In the real world, though, people obviously do not all share the same beliefs; they have *heterogeneous* expectations. By itself, the homogeneous expectations assumption is not important: a version of the CAPM holds where the expected returns are simply averages across the beliefs of all investors.\(^{12}\) But, in combination with the next assumption, heterogeneous expectations can produce significant deviations from the CAPM. 

5. No taxes or transactions costs. 

Taxes matter. Chapter 12 shows that taxes affect expected returns and can be regarded as a systematic factor. Transactions costs, meanwhile, also vary across securities. We should expect that for very illiquid markets with high transactions costs, there may be more deviations from the CAPM. This is indeed the case, and chapter 13 discusses various liquidity premiums in more detail. 

There is another effect of transaction costs when trading frictions are combined with heterogeneous investors. If investors cannot short, then investor beliefs matter. Optimists may prevail in pricing because the pessimists’ beliefs are not impounded into stock prices. Pessimists would like to short but cannot, and so stock prices reflect only the belief of optimists. Thus, investor beliefs become a systematic factor. While there are behavioral versions of this story, the original setting of Miller (1977), where this concept was developed, was a rational setting. Related to this assumption is the next one, since when individuals move prices, markets are likely to be illiquid and there are many trading frictions:
6. Individual investors are price takers. The informed investor is trading and moving prices because he has some knowledge that others do not have. But when these trades are large, they move prices, which leads us to . . .

7. Information is costless and available to all investors. Processing and collecting information is not costless, and certain information is not available to all investors. Information itself can be considered a factor in some economic settings, as in Veldkamp (2011). The CAPM applies in a stylized, efficient market; we should think that additional risk premiums can be collected in more inefficient securities markets, especially where information is very costly and not available to many investors. As we explore in the next few chapters, this is indeed the case; several deviations from the (p.209) CAPM are strongest in stocks that have small market capitalizations and trade in illiquid markets where information is not promulgated efficiently.

In summary, we expect that when the assumptions behind the CAPM are violated, additional risk premiums should manifest themselves. These include macro factors, which should affect investors’ nonfinancial considerations, effects associated with the asymmetric treatment of risk, illiquidity and transactions costs, and taxes. We should expect failures of the CAPM to be most apparent in illiquid, inefficient markets. The assumption, in particular, of perfect information is one of the reasons why modern economists no longer believe that markets are efficient in the form the original CAPM specified.
6. The Fall of Efficient Market Theory
Today, economists do not believe in perfectly efficient markets. In fact, markets cannot be efficient in their pure form. The modern notion of market near-efficiency is developed by Sanford Grossman and Joseph Stiglitz (1980), which forms part of the collection of papers for which Stiglitz was awarded his Nobel Prize in 2001. Grossman and Stiglitz describe a world in which markets are nearly efficient, and in doing so they address a conundrum that arises from the costless information assumption of the CAPM. Suppose that it is costly to collect information and to trade on that information, as it is in the real world. Then, if all information is in the price already, why would anyone ever invest in gathering the information? But if no one invests in gathering the information, how can information be reflected in security prices so that markets are efficient? It is then impossible that markets be efficient in their pure form.

Grossman and Stiglitz develop a model in which markets are near-efficient. Active managers search for pockets of inefficiency, and in doing so cause the market to be almost efficient. In these pockets of inefficiency, active managers earn excess returns as a reward for gathering and acting on costly information. In the assumptions of the CAPM discussed above, we should expect these pockets of inefficiency to lie in market segments that are illiquid, with poor information dissemination and where outsized profits may be hard to collect because trading on these anomalies will likely move prices. Whether active managers actually do earn excess returns for their investors (rather than for themselves) is a topic that we cover extensively in the third part of this book.

(p.210) The near-efficient market of Grossman and Stiglitz fits closely with the multiple factor risk framework of the APT developed by Ross (1976). In Ross’s multifactor model, active managers and arbitrageurs drive the expected return of assets toward a value consistent with an equilibrium trade-off between risk and return. The factors in the APT model are systematic ones, or those that affect the whole economy, that agents wish to hedge against. In their purest form the factors represent risk that cannot be arbitrated away, and investors need to be compensated for bearing this risk.
Despite the modern notion that markets are not perfectly efficient, a large literature continues to test the Efficient Market Hypothesis (EMH). The implication of the EMH is that, to the extent that speculative trading is costly, active management is a loser’s game and investors cannot beat the market. The EMH does give us a very high benchmark: if we are average, we hold the market portfolio and indeed we come out ahead simply because we save on transactions costs. Even if we know the market cannot be perfectly efficient, tests of the EMH are still important because they allow investors to gauge where they may make excess returns. In the Grossman–Stiglitz context, talented investors can identify the pockets of inefficiency where active management efforts are best directed.

The EMH has been refined over the past several decades to rectify many of the original shortcomings of the CAPM including imperfect information and the costs associated with transactions, financing, and agency. Many behavioral biases have the same effect and some frictions are actually modeled as behavioral biases. A summary of EMH tests is given in Ang, Goetzmann, and Schaefer (2011). What is relevant for our discussion is that the deviations from efficiency have two forms: rational and behavioral. For an asset owner, deciding which prevails is important for deciding whether to invest in a particular pocket of inefficiency.

In a rational explanation, high returns compensate for losses during bad times. This is the pricing kernel approach to asset pricing. The key is defining those bad times and deciding whether these are actually bad times for an individual investor. Certain investors, for example, benefit from low economic growth even while the majority of investors find these to be bad periods. In a rational explanation, these risks premiums will not go away—unless there is a total regime change of the entire economy. (These are very rare, and the financial crisis in 2008 and 2009 was certainly not a regime change.) In addition, these risk premiums are scalable and suitable for very large asset owners.

In a behavioral explanation, high expected returns result from agents’ under- or overreaction to news or events. Behavioral biases can also result from the inefficient updating of beliefs or ignoring some information. Perfectly rational investors, who are immune from these biases, should be able to come in with
sufficient capital and remove this mispricing over time. Then it becomes a question of (p.211) how fast an asset owner can invest before all others do the same. A better justification for investment, at least for slow-moving asset owners, is the persistence of a behavioral bias because there are barriers to the entry of capital. Some of these barriers may be structural, like the inability of certain investors to take advantage of this investment opportunity. Regulatory requirements, for example, force some investors to hold certain types of assets, like bonds above a certain credit rating or stocks with market capitalizations above a certain threshold. If there is a structural barrier to entry, then the behavioral bias can be exploited for a long time.

For some risk premiums, the most compelling explanations are rational (as with the volatility risk premium), for some behavioral (e.g., momentum), and for some others a combination of rational and behavioral stories prevails (like value/growth investing). Overall, the investor should not care if the source is rational or behavioral; the more appropriate question is whether she is different from the average investor who is subject to the rational or behavioral constraints and whether the source of returns is expected to persist in the future (at least in the short term). We take up this topic in detail in chapter 14, where I discuss factor investing.

7. The 2008–2009 Financial Crisis Redux
The simultaneously dismal performance of many risky assets during the financial crisis is consistent with an underlying multifactor model in which many asset classes were exposed to the same factors. The financial crisis was the quintessential bad time: volatility was very high, economic growth toward the end of the crisis was low, and there was large uncertainty about government and monetary policy. Liquidity dried up in several markets. The commonality of returns in the face of these factor risks is strong evidence in favor of multifactor models of risk, rather than a rejection of financial risk theory as some critics have claimed. Assets earn risk premiums to compensate for exposure to these underlying risk factors. During bad times, asset returns are low when these factor risks manifest. Over the long run, asset risk premiums are high to compensate for the low returns during bad times.
Some commentators have argued that the events of 2008 demonstrate the failure of diversification. Diversification itself is not dead, but the financial crisis demonstrated that asset class labels can be highly misleading, lulling investors into the belief that they are safely diversified when in fact they aren’t. What matters are the embedded factor risks; assets are bundles of factor risks. We need to understand the factor risks behind assets, just as we look past the names and flavors of the things that we eat to the underlying nutrients to ensure we have enough to sustain us. We take on risk to earn risk premiums in the long run, so we need to understand when and how that factor risk can be realized in the short (p.212) run. Some have criticized the implementation of diversification through mean-variance utility, which assumes correlations between asset classes are constant when in fact correlations tend to increase during bad times.\(^1\) Factor exposures can and do vary through time, giving rise to time-varying correlations—all the more reason to understand the true factor drivers of risk premiums.

Notes:

\(^1\) See Welch (2008) and Graham and Harvey (2001), respectively.

\(^2\) Chapter 17 summarizes the history of index funds.

\(^3\) Technically speaking, the market is the wealth-weighted average across all investors.

\(^4\) A textbook MBA treatment is Bodie, Kane, and Marcus (2011). A more rigorous treatment is Cvitanić and Zapatero (2004).

\(^5\) The constants \(a\) and \(b\) can be derived using equation (6.3) as

\[
\begin{align*}
a &= \frac{1}{\sigma^2_m} \left[ \mu_m \sigma_m + \frac{\mu_m \sigma_m - \tau}{(1+\gamma) \sigma^2_m} \right], \\
b &= -\frac{\mu_m \tau}{(1+\gamma) \sigma^2_m},
\end{align*}
\]

where \(\mu_m = E(r_m)\) and \(\sigma^2_m = \text{var}(r_m)\) are the mean and variance of the market returns, respectively. Note that the coefficient \(b\) multiplying \(m\) is negative: low values of the SDF correspond to bad times, which in the CAPM are given by low returns of the market.

\(^6\) Related to this is that the requirement for the SDF is very weak: it requires only no arbitrage, as shown by Harrison and
Kreps (1979). The CAPM, and other specific forms of $m$, on the other hand, require many additional onerous and often counterfactual assumptions.

(7) And, beyond the scope of this book, there are many useful statistical techniques for estimating $m$ based on statistical “projections” similar to the estimation methods for ordinary least squares regressions based on the notation in equation (6.4).

(8) See, for example, Cochrane (2001) for a straightforward derivation.

(9) There is a third type of factor based solely on statistical principal components, or similar (dynamic) statistical factor estimations of the APT. A pioneering example of these is Connor and Korajczyk (1986). These generally lack economic content, and so I do not discuss them here.

(10) Mayers (1973) is the seminal first reference. See also Constantinides and Duffie (1996), Jagannathan, Kubota, and Takehara (1998), Storesletten, Telmer, and Yaron (2007), and Eiling (2013).

(11) These effects come in two forms. First, there is the risk premium associated with individual stock higher moments. These are properties of each individual stock. See Mitton and Vorkink (2007), Boyer, Mitton, and Vorkink (2010), and Amaya et al. (2012) for skewness risk premiums of this form. Second, there is the risk premium coming from how stock returns covary with higher moments of the aggregate market. Harvey and Siddique (2000), Dittmar (2002), and Chang, Christoffersen, and Jacobs (2013) show that there are risk premiums for co-skewness and co-kurtosis, which result from the co-movement of stock returns with skewness and kurtosis moments of the market portfolio.

(12) See Williams (1977).

(13) The “classical” notions of weak, semi-strong, and strong efficiency were laid out by Fama (1970) and are obsolete. Fama was awarded the Nobel Prize in 2013. In that year, the Nobel Prize committee also gave Robert Shiller the prize, representing the opposite viewpoint of behavioral, or non-rational, influences on financial markets.
See Ellis (1975) for a practitioner perspective.