Equities

Andrew Ang

DOI:10.1093/acprof:oso/9780199959327.003.0008

Abstract and Keywords

Equities have historically exhibited high returns relative to bonds and cash (bills). The equity risk premium is a reward for bearing losses during bad times, which are defined by low consumption growth, disasters, or long-run risks. Equities are a surprisingly poor hedge against inflation. While theory suggests that equity risk premiums are predictable, predictability is hard to detect statistically. Equity volatility, however, is much more forecastable.

Keywords: equity risk premium, time-varying risk aversion, habit, long-run risk, disaster risk, heterogeneous agents, inflation hedging, time-varying risk premium, predictability, volatility timing

Chapter Summary

Equities have historically exhibited high returns relative to bonds and cash (bills). The equity risk premium is a reward for bearing losses during bad times, which are defined by low consumption growth, disasters, or long-run risks. Equities are a surprisingly poor hedge against inflation. While theory suggests that equity risk premiums are predictable,
predictability is hard to detect statistically. Equity volatility, however, is much more forecastable.
1. The Lost Decade
The 2000s were the Lost Decade for stock returns. Figure 8.1 plots the cumulated returns of $1 invested at January 1, 2000 through December 31, 2010 in the S&P 500, Treasury bonds, or Treasury bills (which I call cash). The performance of stocks was dismal. Stock returns started to decline in late 2000 and then the terrorist attacks of September 11, 2001, and the subsequent recession sent stock returns into a tailspin. By September 2002, stocks ended up cumulatively losing more than 40 cents of that initial $1 investment. Stocks recovered over the mid-2000s but lost money again in 2007 as subprime mortgages started to deteriorate. Then came the global financial crisis. In 2008, stocks lost 37% of their value. As policymakers stabilized financial markets and the economy began to improve, stocks bounced back in 2009. But the initial $1 invested in stocks at the beginning of the decade had grown to only $1.05 by December 31, 2010. Investors would have done better simply holding Treasury bills, which cumulated to $1.31 at December 31, 2010. Bonds, however, trounced stocks and bills, finishing the decade at $2.31.

While stocks did poorly in the Lost Decade, they have exhibited a high risk premium relative to bonds and cash over long periods. Figure 8.2 plots average returns (on the y-axis) and volatilities (on the x-axis) for nineteen countries and the world from 1900 to 2010 for stocks and bonds as reported by Dimson, Marsh, and Staunton (2011). The countries are Australia, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, South Africa, Spain, Sweden, Switzerland, the United Kingdom, and the United States. The United States and the world are highlighted separately. Notice in Figure 8.2 how equities around the world congregate toward the upper right-
hand corner of the graph, while bonds around the world are grouped together at the lower left-hand corner. Stocks, therefore, have yielded much higher returns than bonds over long periods of time, but with more volatility. In the United States, the equity return has been 11.4% compared to 5.2% for bonds. Equity return volatility was 20.0% compared to 8.6% for bonds. Thus, over the long run, there has been a considerable equity premium in excess of bonds, but it comes with significantly higher volatility.

Going forward, will the equity premium continue to be high? Or will the next ten years be another Lost Decade for stocks? To answer these questions, we must understand what factors explain equity return performance and volatility and ascertain whether these risk factors will persist in the future. These are issues of how the aggregate market is priced and moves over time.

2. The Equity Risk Premium
In previous chapters we’ve described how risk premiums arising from factor risks compensate investors for bearing losses during bad times. One of those factors, described in chapter 7, was consumption—which was part of a broader set of variables capturing economic growth in general. In a consumption-based asset pricing model, all risk is summarized by consumption—bad times occur when society, or all agents in the economy, are consuming less. We summarize all agents in society by assuming a representative agent, which we can think about as the average person or investor. To compensate investors for holding equities, the average investor earns an equity risk premium. In the most basic consumption models, there is only one factor, per capita real consumption growth, and other factors matter only to the extent they affect
consumption. This makes sense because ultimately we care about what is finally consumed by agents in the economy.

The *equity premium puzzle* is that, using consumption as a risk factor, the equity premium should be very modest. How modest? In 1985, a paper written by Rajnish Mehra and Edward Prescott challenged the profession. They claimed that the equity premium over risk-free assets should be well below 1% for reasonable levels of risk aversion, which are levels of risk aversion between 1 and 10. These are the risk aversion levels that most individuals have (see chapter 2). In contrast, historically the equity premium has been very high. Over 1900–2010, the Dimson, Marsh, and Staunton (2011) data reported in Figure 8.2 show that the average return of U.S. equities in excess of bonds was 6.2% and in excess of bills was 7.4%. The world equity premium in excess of bonds was 5.0% and in excess of bills was 6.0%. Thus, equities have historically had much higher returns than predicted by simple economic models.

Mehra and Prescott titled their paper, “The Equity Premium: A Puzzle,” and it quickly spawned a new literature.¹ They based their model on Lucas (1978), which is one of the most influential papers in financial economics and was a factor in Robert Lucas’s Nobel Prize in 1995. Prescott himself would go on to win the Nobel in 2004 for advancements in macroeconomic modeling.

(p.243) Lucas showed how asset prices responded to an economy’s factors and that these prices were set by agents’ optimal consumption and market clearing. In Mehra and Prescott’s adaptation of Lucas, the factor was consumption, and there was a representative agent with constant relative risk aversion (CRRA) preferences (see chapter 2).² Just as the capital asset pricing model prices all assets through exposure through the market factor, consumption-based asset pricing explains the returns of all assets through exposure to the consumption factor. The relation between the consumption factor, though, and how it enters the risk premiums of individual assets is a transformation that depends on the representative agent’s degree of risk aversion.

To gain intuition why the returns of equity are anomalously high relative to this simple setting, Table 8.3 reports log returns of the S&P 500 with real consumption growth over
June 1947 to December 2010. The Sharpe ratios are approximately 0.3 using either a nominal short rate of 5% or a real short rate of 2%. Table 8.3 shows that the volatilities of equity returns and consumption growth do not line up: equity volatility is high, around 17%, while consumption volatility is low, around 2%. Equity returns are also poorly correlated with real consumption growth, with a correlation of below 15%. Thus, when bad times of low consumption growth occur, equity returns do not tend to fall. Since equities do not exhibit losses when agents experience bad times, as measured by times of slowing consumption, the equilibrium risk premium should be low.

Table 8.3

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>Real Stock</th>
<th>Real Consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stock Returns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>10.4%</td>
<td>7.1%</td>
<td>3.3%</td>
</tr>
<tr>
<td>Stdev</td>
<td>16.5%</td>
<td>16.6%</td>
<td>1.7%</td>
</tr>
<tr>
<td>Sharpe Ratio</td>
<td>0.328</td>
<td>0.308</td>
<td></td>
</tr>
</tbody>
</table>

A skeptical reader might now be asking, “So what?” Consumption does not move around very much, and there are better proxies for economic growth (see chapter 7). I have also told you that most agents do not have mean-variance type utility functions, of which CRRA is an example (see chapter 2). Is this equity (p.244) premium puzzle—that equity returns are high relative to this highly stylized, artificial, Mehra–Prescott economy—purely academic, or does it have relevance for how we invest assets in the real world?

The equity premium puzzle is important because many investors hold large amounts of equities based on the historical record of high returns. Will such returns persist going forward, as in our motivating example, or will equity returns permanently enter a new phase that looks more like an enduring version of the 2000s’ Lost Decade? Understanding the reasons for the equity premium provides a rational basis for deciding whether we should continue to hold lots of
equities in the future, rather than simply continuing to do so because that’s what we have done in the past. While economists have developed many explanations, I cover four that are most relevant for investors.

3. Explanations for the Equity Premium Puzzle
3.1. Market Risk Aversion Is (Sometimes) Very High

The first resolution to the equity premium puzzle is that the market’s risk aversion is extremely high. It can be shown that a lower bound for the risk aversion of the market, $\gamma$, or the representative agent’s risk aversion, is given by

$$\gamma > \frac{\text{Sharpe Ratio}}{\sigma_c},$$

where $\text{Sharpe Ratio}$ denotes the Sharpe Ratio of the equity market portfolio and $\sigma_c$ is the volatility of real consumption growth.\(^3\)

The lower bounds for risk aversion given by the data in Table 8.3 are approximately 20. These are much larger than the upper bounds of “reasonable” risk aversion levels of 10 claimed by Mehra and Prescott and discussed in chapter 2. If we compute the implied risk aversion levels exactly assuming log-normal consumption growth, then risk aversion of the average investor to match the equity premium is above 120. Estimates like these are pervasive throughout the literature beginning with the seminal work of Hansen and Singleton (1983).

Maybe the representative agent’s risk aversion is high, perhaps even above 120. This is problematic for the obvious reason that individuals do not have risk aversion that high. There is a second reason labeled the risk-free rate puzzle due to Philippe Weil (1989). While very high risk aversion can match the equity premiums in data, high risk aversions also raise the risk-free rate in the Mehra-Prescott model. A very risk-averse representative agent would demand a high premium to (p.245) hold equities, but this agent wishes to have extremely smooth consumption paths over time. Risk-free rates need to be very high to induce them not to smooth consumption. It turns out that the question, “why are equity returns so high?” is the flipside of the question, “why are risk-free rates so low?” If we’re over compensated for investing in equities, we’re undercompensated for investing in bonds.
A class of models with *time-varying risk aversion* can explain the equity premium puzzle. Let us illustrate how they work by looking at the model of *habit utility*. We have come across these utility functions before, in chapter 2, where they were used to construct utility of wealth relative to some level, rather than using the utility of wealth itself. In habit utility, the utility of the representative agent depends on consumption relative to past consumption or a *habit level*. A canonical model is by Campbell and Cochrane (1999).

Consider a young college graduate who is accustomed to sleeping on sofas, sharing apartments, and driving second-hand cars. She gets a job as a barista. Then she suddenly gets laid off. This is, of course, a bad time, but it does not hurt that much because she is already used to her frugal lifestyle. Her consumption has not declined by much relative to her habit. Now suppose that upon graduation she lands a high-flying job (on Wall Street, I guess). Her consumption increases because her income is now very high. She moves into her own place with a wonderful view, buys a brand new car, and starts a wardrobe of designer clothing. Her consumption habit has increased. Now when she is suddenly laid off, this is a very bad time. She actually might have more money from some savings than as a barista, but her habitual consumption level is very high, and so the drop in her consumption relative to her habit hurts a lot. Her *marginal utility* is very high because she is used to having high consumption and has to go back to low consumption. In habit models, it is not the consumption decrease per se that matters; what matters instead is the consumption decrease relative to habit.

Habit models allow *local* risk aversion to be very high. When recessions hit, risk aversion can shoot up to very high levels—the risk aversions of 20 or even 100 required by equation (8.1)—but the increase is temporary. In economists’ jargon, marginal utility is very high during recessions as consumption approaches habit; the agent really values that extra $1, and the curvature of the utility function is very steep. Although consumption itself barely falls during recessions, the small reductions in consumption bring agents perilously close to their habits. Thus, agents become very risk averse during bad times, generating high equity risk premiums. Equity prices fall in recessions to generate high future returns.
In good times, consumption is far above habit. Risk aversion is low, and equity premiums are small. In these good times, marginal utility is low, and the representative agent’s utility function is very flat. During booms, equities are expensive, and there is not much room for future appreciation; hence, future expected returns are low. Since expected returns are high during bad times and low during good times, another benefit of time-varying risk aversion models is that they match how equity returns can be predicted over time, a subject that we cover in (p.246) section 5 below. Overall, the very high risk aversion during bad times dominates the low risk aversion during good times, resulting in a high long-run equity risk premium.

If time-varying risk aversion, which increases dramatically during bad times, is responsible for the equity risk premium, then investors should ask the following question before holding large equity stakes: In bad times, say, of low consumption growth in the basic Mehra–Prescott model or bad times more generally in a multifactor context, do you behave in a less risk-averse fashion than the market?

Most investors and the representative investor, by construction, become much more risk averse during recessions: bad times are spooky, so everyone gets scared—but to greater or lesser degrees. The question is whether you can tolerate these bad times better than the average investor. The Norwegian sovereign wealth fund certainly can since it has stable cash flows, no immediate liabilities, and a well-functioning governance structure. Supreme Court Justices are well situated to tolerate bad times since their income stream is risk free (unless they have borrowed money to the hilt). And many private wealth management clients also find bad times for the economy hurt less for them than for the average investor. These special types of investor have lower risk aversions during downside events and can tolerate larger losses during bad times. They can hold more equities. In doing so, they will earn higher average returns.
3.2. Disaster Risk

Disaster models take bad times to the extreme. The bad times are really, really bad: they represent catastrophic declines in consumption. In this explanation, the equity premium is the reward to compensate for rare catastrophes. This story was introduced by Rietz (1988) and immediately dismissed by Mehra and Prescott (1988) as lacking empirical evidence. But the disaster explanation has made a recent roaring comeback beginning with Barro (2008), who examined the data much more thoroughly. Catastrophes are labeled “black swans” by Nassim Taleb (2004) in the popular literature, although he would object to having them treated in a rational paradigm. They cannot be predicted or at least are very hard to predict, and for disasters to account for high equity risk premiums, the disasters must be very bad indeed.

Mehra and Prescott’s original skepticism was not without merit. Looking only at U.S. data, there has not been a large drop in consumption, except for the Great Depression in the 1930s, and even then it was not catastrophic. Looking more broadly at data from other countries and using a long time series, however, Barro and Ursua (2011) document some sizeable drops in consumption and GDP—even for countries that are large and developed today. For example, Barro and Ursua report that Germany’s consumption declined by 41% in 1945 and Japan’s fell by 50% during World War II. The United States and the United Kingdom have been more tranquil, with the largest decline being 16% for the United States in 1921 and 17% for the United Kingdom in 1945. But these are exceptions. For many countries, macroeconomic disasters have been truly calamitous. Russia’s consumption fell 71% in World War I and the Russian Revolution and again by 58% during World War II (poor Russia!). China’s GDP (not consumption) fell 50% from 1936 to 1946, and Turkey’s consumption fell 49% during World War II.

The disaster explanation is also labeled a *peso problem*, which refers to a very low probability event that is not observed in the sample. The U.S. equity premium is high because it reflects the probability of a disaster, but so far we have not observed that disaster. The first published use of the term “peso problem” appears in Krasker (1980) in the context of the phenomenon that the Mexican peso traded at a steep discount on the forward market during the early 1970s. It appeared to be mispriced, but market participants were expecting in a
devaluation that eventually occurred in August 1976. Researchers only looking at the early 1970s sample would have concluded that an asset price—the forward peso—appeared anomalous because a crash had not occurred in the sample available to them. Likewise, the U.S. equity premium may appear too high because we have not observed a disaster in the U.S. sample.

If equities command a risk premium because of infrequent crash risk, then the pertinent question to ask an investor is how he can weather such disasters. Of course, there are some disasters so ruinous that an entire society disappears, and a new one takes its place, such as the Russian Revolution in 1918 and the rise of modern Communist China, where all domestic investors are wiped out. (This by itself should be a good reason for diversifying across countries as mean-variance investing advocates; see chapter 3.) The fact that we have not seen a truly terrible consumption disaster in the United States should give us pause. The extreme advice implied by this theory is that equity investors should stockpile AK-47s and MREs in their basements in case society collapses. Before the cataclysm comes, equity returns will be high; that equity returns are in fact high, on average, foretells the cataclysm to come. We just haven’t seen it yet.

(p.248) My interpretation of the disaster theory for asset owners is that it reemphasizes the importance of downside outcomes, as I stressed in chapter 2. There, I stressed that mean-variance utility does not cut it in capturing how investors respond to bad events. The disaster story says that those bad events can indeed be very bad, and they are systematically very bad for the entire economy. We need to think about how investors behave in these bad outcomes—a joint statement about the utility function and the data-generating process, which must capture disaster events. Investors who are relatively the best off in severe economic crashes (all investors will be affected adversely during disasters; it is a question of how much) are in the best position to hold large amounts of equities. The overall message is that investors need to think about how bad these events can be, how they can respond to meeting their liabilities, and how they react to these risks when these bad events arrive.
Disasters suggest that another explanation for the high equity premium is that it is a result of markets that have survived ex post. In the early 1900s, you could have invested in China or Russia, and you would have lost everything. You could also have invested in Hungary, Czechoslovakia (which were together part of the Austro-Hungarian Empire), Poland, and Greece, but you would not have been able to trade equities again in those countries until the late twentieth century. Jorion and Goetzmann (1999) argue that the realized equity premium is high because we compute the equity return on countries that have survived. Had we computed the returns including all the countries whose markets disappeared, then the premiums would be much lower. Thus, survivorship bias makes the equity premium too high. The true equity premium, which is the one that is relevant for the future, is lower.

I believe this argument only increases the mystery of the equity premium. The equity premium is measured relative to bonds or bills, and sovereign bonds in many countries fared worse than equities. If you invested in German equities in 1900, there were long periods during which markets were closed (like the credit crisis of the early 1930s, World War II, and the occupation by the Allies after World War II), but equity markets survived and eventually German equity holders did well. Holders of German bonds got nothing. First they were ravaged by hyperinflation in the 1920s, and then there was a default on Reichsmark claims when the Deutschmark was introduced in 1948. Of course, in some countries, including Russia and China, both equity holders and debt holders were wiped out. But, in many countries with structural breaks, equity holders eked out a (sometimes modest) long-term return while bond holders were decimated. If the equity premium is measured relative to bonds, this survivorship bias of markets exacerbates the equity premium puzzle. The historical superiority of equities relative to bonds in disasters does carry some implications for factor investing, which I discuss in chapter 14.
3.3. Long-Run Risk

Real consumption growth, which we observe in data, has an autocorrelation close to zero. The data in Table 8.3, for example, exhibit a quarterly autocorrelation of 0.08. The basic Mehra–Prescott model assumes that the consumption shocks are independent and identically distributed (i.i.d.), or that they cannot be forecasted. In an influential article, Bansal and Yaron (2004) change this assumption to specifying that the mean of consumption growth is not constant but wanders around so slowly that it is extremely hard to distinguish between an i.i.d. process and this slow, persistent process where there is a small amount of forecastability. This changing consumption process is what Bansal and Yaron call long-run risk. Bansal and Yaron also alter the basic CRRA preference in the Mehra–Prescott framework so that the representative agent cares much more about long-run risks. The Bansal–Yaron long-run risk model has been adapted to many different contexts and explains a wide variety of stylized facts including the equity premium puzzle and the risk-free rate puzzle of this section.

Europe at the time of writing is a wonderful example of the long-run risk story at play. Current shocks, say, from a particular raucous parliamentary vote in some euro zone country, have an effect on today’s expectation about future economic growth. But they also have an effect on economic growth ten to twenty years in the future. The vote could diminish current productivity, but lowering current productivity might keep Europe on a low-growth path for a long time. Long-run risks exist. The special preferences Bansal and Yaron employ differentiate between the short-run and the long-run effect of these shocks. Equities are particularly sensitive to long-run risk, not surprisingly, because equity is a long-lived (in fact, perpetual) security.

The second channel that Bansal and Yaron employ is that the fundamental consumption factor exhibits time-varying volatility. Agents dislike volatility components and asset prices reflect the volatility risk. As volatility increases, asset valuation declines. The Bansal–Yaron model has separate risk compensations for consumption growth and for consumption volatility.

There are three lessons for an asset owner:
1. Many assets can lose money in the short run. This is the standard channel of how we should evaluate risk.
2. What appears safe in the short run can, in fact, be quite risky over the long run. The market cares about long-run risk, and small adjustments made today can result in large effects in twenty to thirty years time. In the context of factor model theory, assets with large exposures to long-run risk (high long-run risk betas) need to have high returns. It turns out that assets with high long-run risk betas include value stocks and other popular investment strategies such as currency carry trades.\(^{11}\)
3. Volatility as a risk factor matters.
   In Bansal–Yaron, it is consumption volatility, but in a more general setting, it is macro volatility. Volatility risk is different from consumption (macro) risk. It carries its own risk premium. We discussed volatility as a risk factor in chapter 7.

3.4. Heterogeneous Investors
One criticism of Mehra–Prescott involved the assumption of the representative agent. The world has heterogeneous investors, not a single representative agent.\(^{12}\) Perhaps the representative agent isn’t really representative. This was first shown by Jerison (1984), in a paper that never got published but was expanded and extended by Kirman (1992). Jerison and Kirman show that in some circumstances the preferences of the representative agent are not the (weighted) average of preferences of individual agents in the economy. To be blunt, if all agents in the economy prefer bananas over apples, it is possible to construct a representative agent who prefers apples over bananas. Furthermore, how that representative agent feels about apples versus bananas in different states of the world leads to the same prices as the economy where individual agents optimize and prefer bananas over apples. Thus, rejecting a representative agent model with implausibly high risk aversion may not say anything about how a true heterogeneous agent economy works.
The profession remains wedded to representative agent models, though. Some of the reasons are historical and due to tractability; representative agent models are much easier to solve than heterogeneous agent models. But one good reason is that, in a world full of heterogeneous agents, there is still economic meaning for a representative agent even though he may not be the “average” agent, where the average is a simple average or even an average taken over wealth. In fact, some \textit{(p.251)} solution methods of heterogeneous agent models are solved this way by constructing a representative agent with special weights, where the weights do not necessarily correspond to wealth or income.\textsuperscript{13} The weights on some agents may be zero. Agents with small proportions of total wealth in heterogeneous agent models can have very large effects in determining prices.

In this context, the representative agent should not be interpreted as the average agent, but as the \textit{marginal} agent. Prices are set through how the marginal agent responds to small changes in factor shocks, which is different from how the average agent responds. For the asset owner weighing optimal equity allocations, the context of the previous sections—whether bad times for me correspond to bad times for the average agent or whether my risk aversion increases when the risk aversion of the average agent increases—are identical except that we ask the questions in the context of the agents determining prices at the margin. For example, can I absorb losses during bad times more than the asset managers forced to liquidate at fire-sale prices? Am I exposed to the same margin calls as hedge funds?

In heterogeneous agent models, a new factor arises that determines risk premiums. Asset prices now depend on the cross-sectional distribution of agent characteristics. This can be the cross-sectional distribution (across agents) of wealth, beliefs, labor income shocks, or other variables in which investors differ.\textsuperscript{14} In Constantinides and Duffie (1996), agents exhibit heterogeneity by how much they earn. They also cannot completely hedge the risk of job loss (as in the real world). Constantinides and Duffie show that if income inequality increases during recessions (more formally, the cross-sectional variance of idiosyncratic labor income shocks increases), as it does in data, the equity premium will be high. Intuitively, in recessions, the probability of job losses increases and equities drop in value. That is, equities are poor hedges.
for insuring against the risk of losing your job and won't pay off when your boss has fired you and you need to eat. These considerations make equities quite unattractive and mean that equities must exhibit a high return in equilibrium to induce investors to hold them.

The income shocks that workers face in the Constantinides and Duffie world are an example of market incompleteness. This refers to risks that cannot be hedged or removed in aggregate and thus, in effect, serve to increase the total amount of risk that all agents (or the implied representative agent) face. A similar effect can be engineered by constraints, such as those on borrowing. During bad times, some investors may face binding funding constraints that force them to behave in an extremely risk averse way, which manifests in their selling equities. An investment bank, for example, cannot roll over its debt or there are outflows from an asset manager. Both of these situations can similarly raise the aggregate risk aversion of the economy during bad times.

For an asset owner contemplating a large position in equities, heterogeneity implies that she should think not only about how she is different from the market (or the marginal or representative agent) but also what kind of investor she is relative to the full distribution of investors. The behavior of these other investors and how they interact affects asset prices. The main intuition of bad times survives in these heterogeneous agent models—the equity premium is high because some agents (they do not have to be numerous, but when they matter, they matter a lot) find that equity is unattractive for their set of bad times. These may not be bad times for her, and if so, she should hold an above-average allotment of equities.

The problem in applying the insights of heterogeneous agent models to investment policy is that the world is so heterogeneous. It is unclear what dimensions—risk aversion, wealth, loss capacity, leverage capacity, (lack of) liabilities, income, or aversion to downside events—we should emphasize. It is often difficult, if not impossible, to measure cross-sectional characteristics of investor types, which the heterogeneous agent economies predict should be useful candidate factors. Since a representative agent often arises in these models anyway, my advice is to concentrate on the first three explanations of the equity risk premium in terms of
considering how you differ from the market. The implication from heterogeneous agent models is that you may interpret the “market” as the “marginal investor” as well as the “average investor.”

4. Equities and Inflation
4.1. Stocks Are a Bad Inflation Hedge

Equity is a real security, in the sense that it represents a claim on real, productive assets of firms. But it turns out that equities actually are not a real security in the sense that “real” means “inflation adjusted” or the opposite of “nominal.” In fact quite the opposite: equities are bad at hedging inflation risk.\textsuperscript{16}

That may surprise you but take a look at Figure 8.4, which graphs correlations between inflation and stock returns, excess stock returns relative to T-bills (cash), and T-bill returns from 1926 to 2010. Inflation hedging is about the co-movement of assets with inflation, not about long-run average returns. The correlations are computed over various horizons measured in years on the x-axis. I use log returns and log inflation changes so that the long-horizon returns are sums of the short frequency, one-month returns (see the appendix).
Panel A of Figure 8.4 shows that the short-term correlations of stocks with inflation are low—below 10%. Moving out to four to five years, they peak around 30% and then taper off around 20% at the ten-year horizon. T-bills, however, are a wonderful inflation hedge. At short horizons, below one year, the correlations of T-bills with inflation are above 20% and then increase steadily to around 60% at the ten-year horizon. Because raw stock returns are poor inflation hedges, T-bills are good inflation hedges, and excess stock returns subtract T-bills from stock returns, the excess stock return correlations with inflation are lower than the raw stock correlations with inflation. They are around zero below one year, reach a maximum around 10% around years three to four, and then turn negative around year seven. Clearly, stocks are a poor inflation hedge in terms of the way stock returns co-move with inflation, even though stocks have a high average return.\footnote{17}
The true picture is even worse than Panel A of Figure 8.4. Panel A uses Pearson, or classical, correlations, which most of the literature and practitioners focus on. Panel B plots Spearman, or robust, rank correlations. Spearman correlations are robust to outliers, which greatly impact the calculations of simple correlations. Panel B shows that some of the correlations in Panel A are spurious and caused by outliers; the robust correlations with inflation are lower than the simple correlation measures in Panel A. Notably the correlations of excess stock returns are now negative for all horizons and after one year, range between –20% and –10%.

The correlation of stock returns and inflation has also changed over time, as shown in Figure 8.5. Figure 8.5 plots the rolling ten-year classical and robust correlations of excess stock returns and inflation. For some of the 1940s and 1950s, the correlation was positive, but from the 1950s it has been negative. Recently, post-2000, the correlation is barely positive (still below 10%), at least measured by simple correlations, but it remains negative after accounting for outliers. While it has changed sign, the correlation of excess stock returns and inflation has never reached above 20% and in fact has gone below –40%.

T-bills are a good inflation hedge, producing three- to five-year (simple and robust) correlations with inflation close to 0.5 because short-term interest rates directly reflect expected inflation. This is through the Fisher hypothesis and through the actions of monetary policy, which we cover in chapter 9. However, a good inflation hedge does not necessarily mean an overall high nominal or real return; a good inflation hedge is an asset that tends to move together with inflation. In terms of
returns, T-bills just beat inflation. The average return of T-bills (p.255) in the sample is 3.5%, and the long-run average inflation rate is 2.9%. In comparison, equities have an average return of 9.3% and handily beat inflation in the long run. Equity does well—the equity premium is high (see sections 2 and 3). But equity does not co-move with inflation, and this makes equity a bad inflation hedge.

The negative relation of equities and inflation is not particular to the United States. Bekaert and Wang (2010) regress asset returns around the world onto inflation. A good inflation hedge should have an inflation beta of one. Bekaert and Wang find that inflation betas of equities are generally negative for all developed markets, with an average of −0.25. Even when they are positive in certain developed countries, the inflation betas are low and far from one. For example, the North American inflation beta is −0.42 and the EU inflation beta is 0.27. Interestingly, emerging markets do have positive inflation betas close to one, with an average of 1.01.

High inflation is good for those who owe and bad for their creditors. Corporate debt is denominated almost exclusively in nominal terms and high inflation reduces the debt outstanding in real terms. This is beneficial for borrowers at the expense of lenders. Most companies issue debt, which means that the equity market is itself leveraged. So we would expect that high inflation benefits shareholders at the expense of bondholders. This only exacerbates the puzzle.

(p.256) 4.2. Why Are Stocks a Bad Inflation Hedge?
There are a number of reasons why stocks do badly in periods of high inflation. I will examine two rational stories in the context of Ang and Ulrich (2012), where I develop a model where equities are influenced by the same factors driving bond prices and an equity-specific cash flow factor.

First, high inflation reduces future firm profitability. Inflation is negatively associated with real production; see, for example, Fama (1981). Rising inflation reduces profit margins because, while firms can pass through cost increases to consumers continuously, they can only do so in stages. These are called “menu costs” and, empirically, price rigidity is pervasive, as Nakamura and Steinsson (2008) show. This is a cash flow effect.
Second, there is also a discount rate effect. Times of high inflation are bad times when risk is high. This causes expected returns to be high in a factor story. Even in the consumption-only world of Mehra and Prescott, the historical record has seen high inflation episodes occurring with recessions when growth is slow (stagflation). Thus, when inflation is high, expected returns on equity increase, and this cuts equity prices. Thus, equity prices tend to fall with large inflation increases, resulting in a low correlation between realized inflation and realized equity returns. If this were always true, the correlation between inflation and equity should be –1. It is not because there are other factors driving equity prices that are not perfectly correlated with inflation.

The behavioral explanation of equities’ shortcoming as an inflation hedge was proposed by Modigliani and Cohn (1979). According to this account, investors suffer from money illusion, and they discount real dividends using nominal discount rates instead of using real discount rates. Thus when inflation is high, the market’s irrational expectation causes the market to be undervalued relative to the fundamental value. Thus, in times of high inflation, there are realized low returns on the market leading to low correlations of inflation with stock returns.

Under both the rational and behavioral explanations, equity has low returns during high inflation periods and is a bad inflation hedge. Inflation is a risk factor, but inflation causes equities to perform badly, not well. Consequently, any investor forecasting high inflation in the future relative to what the average investor believes is advised to lower holdings of equities.

5. Predicting Equity Risk Premiums
Active investors spend a lot of effort trying to forecast equity returns. But the consensus view in financial theory is that, while equity risk premiums vary over time, movements are hard to predict. Thus, I advise most investors not to time the market, and this is behind my advice in chapter 4 to rebalance back to constant weights or exposures.

(p.257) 5.1. Theory Says Equity Risk Premiums Are Predictable
Consider the Gordon (1963) Dividend Discount Model, which states that the equity price, P, is the present value of future discounted dividends:
\[(8.2)\]
\[P = \frac{D}{E(r) - g},\]
where \(D\) is the expected dividend next period, \(E(r)\) is the discount rate, and \(g\) is the growth rate of dividends. Equation (8.2) is a valuation formula. Holding \(D\) fixed, prices today are high because \(1/(E(r) - g)\) is low: either future expected returns are low, future growth is high, or both. The common wisdom is that high prices occur because growth will be high. For example, at the time of writing, the high share prices of Apple Inc. (AAPL) and Facebook Inc. (FB) suggest that the future growth of these companies will continue. But generally, high prices forecast low growth. (This is the value effect; see chapter 7).

The Gordon model in equation (8.2) can be equivalently expressed as

\[(8.3)\]
\[\frac{D}{P} = E(r) - g,\]
which states that high dividend yields are caused by high future expected returns, low future growth, or both. We could also write

\[(8.4)\]
\[E(r) = \frac{D}{P} + g,\]
which states that the expected return is the dividend yield plus the cash flow growth. Equation (8.4) says that dividend yields should help forecast expected returns, and this has been examined in finance since Dow (1920).

Academics have examined variance decompositions of dividend yields, which are computed by taking variances of both sides of equation (8.3). By doing so we can decompose movements of dividend yields into movements due to expected returns or discount rates, \(E(r)\), or movements due to \(g\), and movements due to the co-movement of both:

\[\text{var} \left( \frac{D}{P} \right) = \text{var} \left( E(r) \right) + \text{var} (g) - 2 \text{cov} (E(r), g).\]

For the discussion below, let’s assume that you can ignore the covariance term so we write:\[18\]

\[(8.5)\]
\[\text{var} \left( \frac{D}{P} \right) \approx \text{var} \left( E(r) \right) + \text{var} (g).\]
It seems obvious that dividend yields move over time. Figure 8.6 plots dividend yields and earnings yields of the S&P 500 from January 1900 to December 2011. Both series move in tandem, with the correlation being 73%. Since price appears in the denominator of dividend yields and earnings yields, the yield measures tend to be high when the equity market is in the doldrums, like during the Great Depression, the early 1950s, the late 1970s and early 1980s, and 2001. In 2008 when equities fell during the financial crisis, dividend yields were very low. But, during this period earnings yields moved in the opposite direction because earnings contracted a lot, but dividend payouts were sticky.

“What drives dividend yield variation?” turns out to be the same question as “what predicts returns?” Suppose cash flows are unpredictable, or i.i.d.. Then in equation (8.4), expected returns are predictable by the dividend yield: when dividend yields are high or prices are low, future expected returns are high. Equation (8.5) says that all dividend yield movements are driven by discount rate variation. Now suppose everything comes from cash flows and discount rates are constant. Equation (8.4) says that expected returns would change over time if we can forecast the future cash flows of companies, and equation (8.5) says that all variation in dividend yields comes from changing cash flows.

Do dividend yields move around because discount rates move, cash flows move, or both? Academic opinion is divided. John Cochrane, in his presidential address to the American Finance Association in 2011, gave a one-sided view and stated that all variation of dividend yields comes from expected returns, not cash flows. This used to be the prevailing opinion in finance due to Campbell (1991) and Cochrane (1992), even though contemporaneous dissenters argued that returns were not predictable (e.g., Goetzmann and Jorion in 1993). The
Campbell–Cochrane view is that dividends are i.i.d. and dividend yields predict expected returns, not future cash flows. In the Bansal and Yaron (2004) framework (p.259) to explain the equity premium, however, it is exactly the opposite. In Bansal and Yaron’s world, almost all dividend yield variation comes from cash flows, and there is essentially no discount rate channel. The risk premiums are still predictable according to equation (8.4), but they are driven only by predictable cash flows.

I have a stake in this literature and, unlike such “true believers” in discount rate predictability as Cochrane, my research documents the existence of cash flow predictability. In Ang and Liu (2007), I show that since both dividend yields and equity return volatility vary over time (see also section 6 below), it must be the case that risk premiums also exhibit predictable components. Overall, my opinion is that the truth lies somewhere in between the Cochrane and the Bansal–Yaron extremes: there is both cash flow and discount rate predictability, but it is hard to detect both (see below).

What an investor should take away is that dividend yields vary over time, and, as a result, expected returns and cash flows are predictable. In the discount rate story, low prices mean high returns in the future and times of low prices are good times to buy equities. If we can forecast high future cash flows according to the cash flow story, we want to hold equities. A smart investor might be able to use these facts to her advantage.

5.2. Theory Says Time-Varying Equity Risk Premiums Are Hard to Estimate

Theory says equity risk premiums are predictable, but theory also says that the amount of predictability is small. Thus, it is hard to statistically detect predictability in data and even harder to trade on it.
Predictability Regressions

A common statistical model to capture predictability is a predictability regression:

\[(8.6)\]

\[r_{t+1} = c + b \cdot X_t + \varepsilon_t,\]

where \(r_{t+1}\) is the market excess return and \(X_t\) is a set of predictive variables. The predictors range from the economically intuitive factors, like macro factors, to more esoteric variables, like sentiment-based factors based on Twitter or Facebook.\(^{21}\) In regression (8.6), the strength of predictability is measured by the statistical significance of the predictive coefficient, \(b\). Econometricians like to see high \(t\)-statistics, or low \(p\)-values, with a standard cut-off of a \(p\)-value of 5\% called “statistically significant.” If the returns and the predictive instruments have been standardized, that is, set to zero mean and a variance of one, then \(b\) represents a correlation between a predictive variable today and next-period returns. In this case, the regression \(R^2\) is simply given by squaring the correlation coefficient, \(b^2\).

We are also interested in long-horizon predictability, which we can measure statistically by extending the regression (8.6) to multiple periods of returns on the left-hand side. We use log returns so we can add them. For example, for a three-period excess return we have

\[(8.7)\]

\[r_{t+3} + r_{t+2} + r_{t+1} = c + b \cdot X_t + \varepsilon_{t+3},\]

Because we have limited data, we generally run long-horizon regressions using overlapping data. That is, when we consider the timing of the variables at time \(t+1\) predicting returns over the next three periods, we predict the one-period excess returns from \(t+1\) to \(t+4\), \(r_{t+4}\), \(r_{t+3}\), and \(r_{t+2}\). Thus, considering the two regressions with right-hand side variables at \(t\) and \(t+1\), we overlap the one-period returns \(r_{t+3}\) and \(r_{t+2}\) on the left-hand side. This overlap is behind the complicated notation, \(\varepsilon_{t+3}\), in the residuals of (8.7), which denotes that the residuals are realized at \(t+3\), but they involve returns over the last three periods (the second term in the subscript). The overlapping data turn out to induce some very nasty statistical properties that cause regular ordinary least squares statistical inference to highly overstate how much true predictability there is in
data. This is not just a problem for geeky econometricians; it has large implications for investors, as we shall see.

How much predictability should we expect? Not much. Ross (2005, 2012) and Zhou (2010) show that the $R^2$ in the predictive regression (8.6) should be very low. Technically, Ross and Zhou show that the regression $R^2$ is bounded by the variance of the pricing kernel (see chapter 6), and the factors entering the pricing kernel, like consumption, are generally not volatile.\textsuperscript{22} Ross computes a bound of 8% and Zhou’s bound is even lower. We should expect $R^2$’s to be low and typically below 5%. That is, 95% of the movements of the market should be unpredictable, and any strategy that claims to predict future returns with high $R^2$’s should be viewed with great suspicion. The low degree of predictability is consistent with the near-efficient markets of Grossman and Stiglitz (1980) in chapter 6: profitable market-timing strategies are rare and statistically hard to detect.

The data bear this out. Empirically, it is hard to find robust statistical evidence of predictability. Table 8.7 reports correlation coefficients of next-period returns on various predictors at the beginning of the period. The base unit of measurement is one-quarter, so the one-period ahead return in regression in (8.6) is over the next quarter. I also run long-horizon regressions of the form (8.7) over one year, two years, and five years. These are also reported as correlation coefficients. Table 8.7 reports robust $t$-statistics that account for the overlapping observation (p.261) (p.262) (p.263) problem and the problem of time-varying volatility, which also causes regular ordinary least squares to overreject the null of predictability too often.\textsuperscript{23} Table 8.7 considers a variety of variables used in the literature ranging from macro variables, interest rates and spreads, and past excess returns.
### Table 8.7

<table>
<thead>
<tr>
<th></th>
<th>1 Qtr</th>
<th>1 Year</th>
<th>2 Years</th>
<th>5 Years</th>
<th>Sample</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dividend Yield</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jan 1926 - Dec 2011</td>
<td>CRSP</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.12</td>
<td>0.23</td>
<td>0.33</td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(1.12)</td>
<td>(1.25)</td>
<td>(1.48)</td>
<td>(1.95)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td><strong>0.18</strong></td>
<td><strong>0.33</strong></td>
<td><strong>0.41</strong></td>
<td><strong>0.48</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(2.23)</td>
<td>(2.55)</td>
<td>(2.79)</td>
<td>(2.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volatility VIX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Jan 1986 - Dec 2011</td>
<td>CBOE</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.12</td>
<td>0.10</td>
<td>0.13</td>
<td>-0.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(0.96)</td>
<td>(0.63)</td>
<td>(0.75)</td>
<td>(-0.47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past Volatility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sep 1963 - Dec 2011</td>
<td>CRSP</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.05</td>
<td>0.09</td>
<td>0.13</td>
<td>0.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(0.55)</td>
<td>(0.65)</td>
<td>(0.78)</td>
<td>(0.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Over Last Quarter]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dec 1926 - Dec 2011</td>
<td>CRSP</td>
</tr>
<tr>
<td>Lagged 1-Year Return</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.01</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(-0.12)</td>
<td>(-0.26)</td>
<td>(-0.55)</td>
<td>(-1.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Past 12 Mths to Past 1 Mth]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sep 1926 - Dec 2011</td>
<td>Ibbotson</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.07</td>
<td>-0.12</td>
<td>-0.15</td>
<td>-0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(-0.12)</td>
<td>(-0.26)</td>
<td>(-0.55)</td>
<td>(-1.39)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Equities

<table>
<thead>
<tr>
<th></th>
<th>1 Qtr</th>
<th>1 Year</th>
<th>2 Years</th>
<th>5 Years</th>
<th>Sample</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>t-stat</strong></td>
<td>(-1.23)</td>
<td>(-1.17)</td>
<td>(-1.05)</td>
<td>(-0.97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consumption-Wealth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.16</td>
<td>0.30</td>
<td>0.42</td>
<td>0.53</td>
<td>Mar 1953 - Dec 2011</td>
<td>Lettau and Ludvigson</td>
</tr>
<tr>
<td>Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>(2.92)</td>
<td>(2.62)</td>
<td>(2.23)</td>
<td>(1.34)</td>
<td>(2001a) updated</td>
<td></td>
</tr>
<tr>
<td><strong>Term Spread</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.11</td>
<td>0.23</td>
<td>0.27</td>
<td>0.35</td>
<td>Mar 1953 - Dec 2011</td>
<td>Federal Reserve Bank</td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>(1.60)</td>
<td>(1.84)</td>
<td>(1.66)</td>
<td>(1.61)</td>
<td></td>
<td>of St Louis</td>
</tr>
<tr>
<td><strong>Credit Spread</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.02</td>
<td>0.08</td>
<td>0.12</td>
<td>0.20</td>
<td>Dec 1918 - Dec 2011</td>
<td>Federal Reserve Bank</td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>(0.15)</td>
<td>(0.34)</td>
<td>(0.44)</td>
<td>(0.67)</td>
<td></td>
<td>of St Louis</td>
</tr>
<tr>
<td><strong>GDP Growth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>0.01</td>
<td>-0.13</td>
<td>-0.11</td>
<td>-0.08</td>
<td>Jun 1947 - Dec 2011</td>
<td>BEA</td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>(0.17)</td>
<td>(-1.37)</td>
<td>(-1.25)</td>
<td>(-0.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Inflation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.07</td>
<td>-0.12</td>
<td>-0.10</td>
<td>-0.03</td>
<td>Mar 1914 - Dec 2011</td>
<td>Federal Reserve Bank</td>
</tr>
<tr>
<td><strong>t-stat</strong></td>
<td>(-1.07)</td>
<td>(-0.94)</td>
<td>(-0.58)</td>
<td>(-0.09)</td>
<td></td>
<td>of St Louis</td>
</tr>
<tr>
<td></td>
<td>1 Qtr</td>
<td>1 Year</td>
<td>2 Years</td>
<td>5 Years</td>
<td>Sample</td>
<td>Source</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>----------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>Industrial Production</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dec 1919 - Dec 2011</td>
<td>Federal Reserve Bank</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.01</td>
<td>-0.04</td>
<td>0.02</td>
<td>-0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Past 1 Year] t-stat</td>
<td>(-0.06)</td>
<td>(-0.25)</td>
<td>(0.08)</td>
<td>(-0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Oil Price</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dec 1945 - Dec 2011</td>
<td>Federal Reserve Bank</td>
</tr>
<tr>
<td>Correlation</td>
<td>-0.10</td>
<td>-0.18</td>
<td>-0.16</td>
<td>-0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Past 1-Year Change] t-stat</td>
<td>(-1.20)</td>
<td>(-1.20)</td>
<td>(-1.08)</td>
<td>(-0.78)</td>
<td></td>
<td>of St Louis</td>
</tr>
<tr>
<td><strong>Unemployment Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dec 1947 - Dec 2011</td>
<td>Bureau of Labor</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.12</td>
<td>0.22</td>
<td>0.19</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-stat</td>
<td>(1.86)</td>
<td>(1.85)</td>
<td>(1.08)</td>
<td>(0.75)</td>
<td></td>
<td>Statistics</td>
</tr>
</tbody>
</table>
Table 8.7 shows that it is very hard to predict returns. I highlight in bold in Table 8.7 all correlation coefficients that are significant at the 95% level—that is, they have p-values smaller than 0.05 (or equivalently they have t-statistics greater than 1.96 in absolute value). There are only two variables that have statistically significant correlations at the 95% level: Shiller’s (2000) cyclically adjusted ten-year earnings yield and the deviations from long-run consumption-wealth trends constructed by Lettau and Ludvigson (2001a) known as “cay.” The latter has look-ahead bias (it is constructed using information not available to investors at the time), so should be interpreted with caution because it cannot be used in an investable strategy.24 The dividend yield has borderline predictive ability at the five-year horizon, with a t-statistic of 1.95. Thus, it is very hard to predict returns, just as Ross and Zhou say it should be.

The Shiller inverse price–earnings ratio that does so well in Table 8.7 and the borderline dividend yield at the five-year horizon is consistent with the Gordon model intuition in equation (8.4). High yields mean low prices. Low prices result from future cash flows being discounted at high expected returns. Consequently, the low price today forecasts that future returns tend to be high. Note that the only significant variables in Table 8.7 are these valuation ratios. Other predictors are documented in an extensive academic literature over different time horizons and samples.25 Overall, it is hard to predict returns.

**Be Wary of Long-Horizon R²s**

Many commentators argue that there is substantial evidence of long-horizon predictability in data. Table 8.7 does not show this. It is there, but, again, it is hard to detect. You cannot conduct statistical inference at long horizons by using standard R² measures. The overlapping observations in equation (8.7) cause long-horizon R²s to be grossly inflated; the simple R²s appear to indicate substantial predictability, but it isn’t actually there. For example, take the dividend yield correlation of 49% with excess returns over the next five years. This implies an R² = (0.49)² = 24%. This looks enormous, especially given that financial theory predicts that the regression R²s should be close to zero.
Investors cannot rely on long-horizon $R^2$'s because they are spurious. This means that while the true predictability is very weak (we say that there is a small degree of predictability in population), the regression $R^2$'s are much larger in small samples. Small samples here can mean hundreds or thousands of years. The intuition for this spurious effect is that the overlapping observations in equation (8.7) artificially induce dependence—the three-period long-horizon return from $t$ to $t+3$ overlaps two periods with the three-period long-horizon return from $t+1$ to $t+4$—and the problem gets worse as the horizons get longer.\textsuperscript{26} Valkanov (2003) and Boudoukh, Richardson, and Whitelaw (2008), among others, derive the correct small sample $R^2$ distributions and show they are hugely biased upward; using the simple $R^2$ always overstates, often grossly so, long-horizon predictability. The robust $t$-statistic corrections get the calculations right; $R^2$'s do not. The moral of the story is that there is not that much predictability in the data—which is as it should be according to theory—and it is also hard to detect at long horizons.

**Predictability Comes and Goes**

A final twist makes it even harder for investors to predict returns in practice: the predictive coefficients themselves (the $b$ parameter in equation (8.7)) vary over time. Statistical models that can accommodate this parameter instability do so by allowing the coefficients to undergo structural breaks or to change slowly over time.\textsuperscript{27} Henkel, Martin, and Nardari (2011) capture regime-dependent predictability. They find that predictability is weak during business cycle expansions but very strong during recessions. Thus, predictability is countercyclical and is observed most during regimes of economic slowdowns.\textsuperscript{28}

**Implications for Investors**

There are time-varying risk premiums, but they are difficult to estimate. If you attempt to take advantage of them, do the following:

1. Use good statistical techniques.
Overstating statistical significance, for example, by using the wrong $t$-statistics and thereby making predictability look “too good,” will hurt you when you implement investment strategies. One manifestation of spuriously high $R^2$ in fitted in samples is that the performance deteriorates markedly (p.265) going out of sample. Consistent with the spurious high $R^2$s, Welch and Goyal (2008) find that the historical average of excess stock returns forecasts better than almost all predictive variables. Use smart econometric techniques that combine a lot of information, but be careful about data mining and take into account the possibility of shifts in regime.29

2. Use economic models.
Notice that the best predictors in Table 8.7 were valuation ratios. Prediction of equity risk premiums is the same as prediction of economic value. If you can impose economic structure, do it. Campbell and Thompson (2008), among others, find that imposing economic intuition and constraints from economic models helps.30

If you’re trying to time the market, then have humility. Predicting returns is hard. Since it is difficult to statistically detect predictability, it will also be easy to delude yourself into thinking you are the greatest manager in the world because of a lucky streak (this is self-attribution bias), and this overconfidence will really hurt when the luck runs out. You will also need the right governance structure to withstand painful periods that may extend for years (see chapter 15). Asset owners are warned that there are very few managers who have skill, especially among those who think they have skill (see also chapters 17 and 18 on hedge funds and private equity, respectively).

Since the predictability in data is weak, investors are well served by taking the i.i.d. environment as a base case. This means dynamic asset allocation with rebalancing to constant weights (or exposures) along the lines of chapter 4. If there is some mean reversion of returns in the data, as suggested by the low returns that follow high prices, and vice versa, then rebalancing back to constant weights will be advantageous. Recall that rebalancing is a counter-cyclical strategy; it buys
when equities have fallen in price, so the equity position is increased when expected returns are high and sells when equities have risen in price, so the equity position is reduced when expected returns are low. You can do better than this if you can predict the future, but you probably can’t. At least rebalancing will get you going in the right direction—the truly optimal (ideal) strategy will only be more aggressive than rebalancing. If you can’t do constant rebalancing, then you are highly unlikely to be able to undertake the optimal investment strategy when the true data-generating process of the equity risky premium exhibits predictability.

(p.266) 6. Time-Varying Volatility
In contrast to the weak predictability of equity risk premiums, there is a great deal of predictability in aggregate market volatilities.31

First, the volatility of equities is relatively high and much higher than fundamental volatilities of dividends (and also of consumption as a factor, as mentioned in section 2). This is labeled “excess volatility” by Robert Shiller (1981). From January 1935 to December 2010, the standard deviation of log returns is 16.0%. In contrast, the standard deviation of dividend growth is 9.6%, and thus return volatility is higher than fundamental volatility. However, taking earnings as fundamentals gives a different story. The volatility of earnings growth is 36.0%, and there is no excess volatility puzzle. But earnings cannot be directly used to value equity because they are generated by a firm owned by both debt and equity holders (it is also an accounting construct and not an actual cash flow payment).32 Dividends accrue only to stockholders. The fact that equity return volatility is higher than dividend volatility is a reflection that some predictability of risk premiums is coming from discount rates and not from cash flows in the Cochrane versus Bansal–Yaron debate.

Models to predict volatility were created by Robert Engle (1982) and are called ARCH models (for “autoregressive conditional heteroskedasticity”). This is a mouthful and means that conditional volatility changes over time (“heteroskedasticity” is Greek for “different dispersions”) in such a way that it is mean-reverting (“autoregressive”). The model was extended by Bollerslev (1986) to GARCH, for “generalized autoregressive conditional heteroskedasticity,”
and that is the name used by industry and the literature. The GARCH model revolutionized volatility modeling, especially for risk management. Engle was awarded the Nobel Prize in 2003.

In the basic GARCH model, conditional variance, $\sigma_t^2$, follows:

\begin{equation}
\sigma_t^2 = a + b\sigma_{t-1}^2 + c\epsilon_{t-1}^2,
\end{equation}

The main effect in equation (8.8) is that volatility depends on itself and is thus persistent (through $b$, which is close to one), or autoregressive, and it is also affected by past shocks (through $c$). Thus, large shocks—like those in 1987 (stock market crash), 1998 (Russian default and emerging markets crisis), and 2008 (p.267) (Lehman default)—lead to higher future volatility. Once these hit, volatility remains high for a certain period of time before mean-reverting. This allows the GARCH model to capture periods of turbulence and quiet. We see this in Figure 8.8, which plots GARCH forecasted volatility with realized volatility (computed using realized daily returns over the month), both at the monthly frequency, from January 1990 to December 2012. The correlation of GARCH volatility at the beginning of the month with realized volatility over the next month is 63%. Compare this correlation with the barely 5% correlations that we were expecting to predict future equity risk premiums! Volatility is quite predictable in equity markets.

If volatility is so predictable, then volatility trading should lead to terrific investment gains. It does. Despite my pessimism on predicting expected returns of the previous section, I am far more enthusiastic on strategies predicting volatilities.
Panel A of Figure 8.9 contrasts cumulated returns from January 1986 to December 2011 of a static 60% equities–40% T-bill strategy with a volatility timing strategy based on VIX, which is also overlaid on the graph. The static strategy is the mean-variance portfolio weight given in equation (2.10) in chapter 2, which for convenience is repeated here:

\[
(8.9) \quad w = \frac{\mu_r - r_f}{\gamma \sigma^2},
\]

where the risk aversion, \( \gamma \), is calibrated to produce a 60%/40% portfolio over the whole sample. The volatility timing strategy replaces market volatility, \( \sigma \), in the denominator of equation (8.9) with the VIX, so that it becomes time varying. (p.268) I hold the numerator equal to the mean excess return over T-bills, \( \mu_r - r_f \), over the whole sample just as in the 60%-40% strategy. That is, the volatility timing strategy only changes volatility and holds the mean fixed. In Figure 8.9, I use the known VIX at the beginning of the current month. There are plenty of examples of volatility strategies, and important references include Fleming, Kirby, and Ostdiek (2001) and Kirby and Ostdiek (2012). The most sophisticated versions of volatility forecasts combine information from past returns using GARCH-type models, models estimated on realized volatilities, and stochastic volatility models estimated from option prices (see, e.g., Andersen et al. (2006)).
Panel A of Figure 8.9 shows that the cumulated returns (left-hand axis) of the volatility timing and the static 60%/40% strategy. I have scaled the returns of both so that they have a target volatility of 10%. The figure shows that the volatility strategy decisively beat the static strategy and was less prone to drawdowns during the early 2000s and the 2008 financial crisis. During these periods VIX (right-hand axis) was high and the volatility timing strategy shifted into T-bills. It thus partly avoided the low returns occurring when volatility spiked. The cumulated return at the end of the sample is 3.06 for the volatility timing strategy compare to 2.14 for the static 60%/40% strategy. Volatility strategies perform strongly.

An obvious question: if volatility timing does so well, why doesn’t everyone do it? Volatility timing requires nimbleness. Panel B of Figure 8.9 shows that the portfolio weights move all over the place. Portfolio weights in equities range from more than 1.5 to nearly zero (as during 2008). Most large investors would have trouble moving in and out of stocks so rapidly or even managing large equities futures positions if done synthetically. So this is not a strategy for all investors.

A deeper question is: what beliefs are implied by the average investor, the representative agent, who holds 100% in the market portfolio and cannot flit in and out of equities by definition? If we rearrange equation (8.9) for the
representative agent who holds $w=1$ in the market portfolio, then we obtain the capital market line of chapter 6 (see equation (6.1)), repeated here:

\[(8.10)\]

$$\mu - r_t = \gamma \sigma^2,$$

and high volatilities should correspond to very high future returns (but be accompanied by low contemporaneous returns). We see little predictive power of volatilities in Table 8.7 and overall little evidence of predictability. What the portfolio choice of equation (8.9) is missing, which was used to derive equation (8.10) under the assumption that the representative agent holds 100% of the market, is terms representing liabilities and the long-horizon investment opportunities of chapter 4. Grouping both of these into “hedging terms,” we have

\[(8.11)\]

$$w = \frac{1}{\frac{1}{\gamma} + \frac{1}{\sigma^2} + \text{hedging terms}},$$

and the hedging terms must be playing very important roles for the representative agent. Guo and Whitelaw (2006) argue that the hedging terms completely dominate the simple linear risk–return relation and offset the effects of volatility.

A closing comment is that the modern practice of risk management almost exclusively measures risk in (souped-up versions of) GARCH models. I believe risk management is the flipside of expected returns. That is, volatility must be linked with risk premiums, and together they determine valuations. This arises naturally (p.270) in economic models and is the point made by Ang and Liu (2007). The next generation of risk management models should embed valuation metrics. When volatility spikes, prices fall. But when prices are low, discount rates tend to be high, and this predicts that returns going forward will be high (see equation (8.4)). We see little direct relation between volatility and future returns (see Table 8.7) because of the hedging terms in equation (8.11) and the fact that the relation between volatility and the risk premium is nonlinear (see also chapter 6). When volatility is high and prices are low, risk is high as measured by high volatility, but these are times of high risk premiums. In terms of valuations, these are times of low risk and opportunities to buy. Of course, one of the main messages from this chapter is that forecasting the risk premium is an inherently difficult exercise.
7. The Lost Decade Redux
The Lost Decade of the 2000s was certainly brutal for stock returns relative to bonds and even T-bills (cash). But over much longer time periods, there has been a considerable equity risk premium relative to bonds and bills. While the equity risk premium is high, equity returns are volatile—and even excessively volatile relative to fundamentals like consumption, dividend, and other macro volatility measures. So low returns over a decade are not unexpected. In fact, it is precisely because equities can perform poorly over extended periods and over bad times in general that equities earn a risk premium. These bad times include periods of low growth, disasters, long-run risks, and high inflation, among other factor risks. I expect the equity premium to be high going forward but over the long run. Over the short run, investors holding equities should steel themselves for the possibility of poor returns. You earn high returns over the long run to compensate for the possibility of poor performance during bad times in the short run.

The equity risk premium is predictable. But theory and empirical evidence suggest that the predictability is hard to detect. Investors should ignore the weak predictability in data and rebalance to constant weights. If they do wish to exploit time-varying risk premiums, then they should be careful to use robust statistical inference, economic models, and restrictions where possible and do it in a framework of good governance structures. The constant rebalancing strategy is naturally countercyclical, and any strategy that takes advantage of mean-reverting risk premiums will use a more aggressive strategy than simply maintaining constant weights.

Notes:
(1) A recent summary is Mehra (2006).

(2) In the context of factor theory (see chapter 6), Mehra–Prescott’s world has a nonlinear pricing kernel, which is given by how the representative agent responds (at the margin) to changes in real consumption growth.

(3) See Cochrane (2001) for a derivation. This is a special case of a more general bound on the pricing kernel (see chapter 6) developed by Hansen and Jagannathan (1991). Lars Hansen received his Nobel Prize in 2013.
More generally, they do not need to be large, sudden negative jumps per se. They can be a series of persistent, large losses in the left-hand tails. Martin (2012) shows that asset prices reflect the possibility of extraordinary bad news that nests but does not necessarily need the assumption of aggregate disaster risk.

The recent financial crisis cannot be considered as an example of an averted disaster to explain the equity premium. While there was a downturn in economic growth and consumption, the downturn happened as a result of the crash in financial prices, rather than the other way round. In disaster models, sharp contractions of consumption happen contemporaneously with falls in asset prices.

For the entire economy. Certainly it was catastrophic for certain individuals like the Joad family, in John Steinbeck’s *Grapes of Wrath*.

Bansal and Yaron (2004) also specify that the volatility of the conditional mean of consumption varies over time.

Predictability in consumption growth had been considered before, but allowing for simple autocorrelated consumption growth actually makes the equity premium puzzle worse with CRRA utility because it smooths out consumption, requiring even higher degrees of risk aversion to match the data. See Dunn and Singleton (1986) for an early attempt.

Specifically, the representative agent wishes to have smooth consumption paths, but the way that these consumption paths are smoothed is different from risk aversion, which measures how an agent treats different risky betas at a given moment in time. These are called Epstein and Zin (1989) preferences, and they are also used in the disaster framework.

See Bansal (2007) for a summary of this literature that was already large by then.

See Bansal, Dittmar, and Lundblad (2005) and Bansal and Shaliastovich (2010), respectively.

Heterogeneous agent models were first examined by Bewley (1977).
These are called Negishi (1960) weights.

Models with heterogeneity in risk aversion, which give rise to the factor of a cross-sectional wealth distribution, were developed by Dumas (1989) and Wang (1996). For models with heterogeneity in beliefs, see David (2008).


An early reference on the negative relationship between equity returns and inflation is Lintner (1975).

Ang, Brière, and Signori (2012) also find that within the cross section of equities, it is difficult to find individual stocks that can hedge inflation risk by computing stock inflation betas. Konchitchki (2011) finds more success in finding individual stocks that correlate highly with inflation using accounting data.

Serious studies do not ignore the covariance term. They assign the covariance term to the other variance terms through orthogonalization (usually Cholesky) procedures.

There is a discount rate channel in Bansal and Yaron because there is stochastic volatility and the risk-free rate is time varying. But the variation in these components is small.

See Ang and Bekaert (2007), Ang (2012b), and Ang and Zhang (2012).

See Bollen, Mao, and Zeng (2011) and Karabulut (2011), respectively.

Denoting the pricing kernel as $m$, Ross (2005) shows that the regression $R^2$ is bounded by $R^2 \leq (1 + r_f)^2 \text{var}(m)$. Zhou (2010) derives a more restrictive bound of $R^2 \leq \rho_{zm}^2 (1 + r_f)^2 \text{var}(m)$, where $\rho_{zm}$ is the multiple correlation between the predictive variables $z$ and the pricing kernel $m$.

The $t$-statistics in Table 8.7 are computed using Hodrick (1992) standard errors, which Ang and Bekaert (2007) show have the correct statistical properties and do not overreject the null of predictability. There are many versions of robust standard errors that still are not “robust” enough, including
the commonly used Hansen and Hodrick (1980) and Newey and West (1987) standard errors, in the sense that they do not have correct size properties.

(24) See Brennan and Xia (2005).

(25) See a summary article by Rapach and Zhou (2011), who show that combinations of variables can do a better job than individual variables at forecasting equity returns.

(26) The literature on spurious regressions calls these partially aggregated variables.

(27) See, for example, Paye and Timmermann (2006) and Lettau and Van Nieuwerburgh (2008) for structural breaks in predictability relations. Dangl and Halling (2012) allow the predictive coefficient to change slowly following a random walk specification. The coefficients of predictability change in Johannes, Korteweg, and Polson (2011) as investors learn over time.

(28) For a summary of regime changes and financial markets, see Ang and Timmermann (2012).


(30) Ang and Piazzesi (2003) impose no-arbitrage constraints and find these improve forecasts of interest rates. Campbell and Thompson (2008) impose that forecasts of equity risk premiums are positive by truncating at zero. A more general approach is Pettenuzzo, Timmermann, and Valkanov (2013), who impose bounds on Sharpe ratios.

(31) Most of this predictability in equity volatility comes from equity volatility predicting itself. There is little predictive ability of macro variables to forecast market volatility, as Paye (2012) shows.

(32) There are many earnings models, but one seminal paper that links the dividend discount model with what are now called residual income models is Miller and Modigliani (1961). These are the same Miller and Modigliani who received the Nobel prize in 1985 for (the irrelevance of) capital structure. My paper on this is Ang and Liu (2001).
There have been many extensions in the literature for different past volatility dependence, asymmetry, including jumps, and so on.