Perspectives on the Equity Risk Premium

Jeremy J. Siegel

The equity risk premium, or the difference between the expected returns on stocks and on risk-free assets, has commanded the attention of both professional economists and investment practitioners for many decades. In the past 20 years, more than 320 articles, enough to fill some 40 economics and finance journals, have been published with the words "equity premium" in the title.

The intense interest in the magnitude of the premium is not surprising. The difference between the return on stocks and the return on bonds is critical not only for asset allocation but also for wealth projections for individual investors, foundations, and endowments. One of the most asked questions by investors is: How much more can I expect to earn from shifting from bonds to stocks?

Academic interest in the equity premium surged after Mehra and Prescott published a seminal article in 1985 titled "The Equity Premium: A Puzzle." By examining the behavior of the stock market and aggregate consumption, they showed that the equity risk premium, under the usual assumptions about investor behavior toward risk, should be much lower than had been calculated from the historical data. Indeed, Mehra and Prescott stated that the equity premium in the U.S. markets should be, at most, 0.35 percent instead of the approximately 6 percent premium computed from data going back to 1872.

The Mehra–Prescott research raised the following question: Have investors been demanding—and receiving—"too high" a return for holding stocks based on the fundamental uncertainty in the economy, or are the models that economists use to describe investor behavior fundamentally flawed? If the returns have been too high, then analysts can justify increased asset allocation to equities and reduced allocation to bonds; if the models are flawed, economists need to develop new models to describe investor behavior.

My discussion of the equity risk premium will be divided into three parts: (1) a summary of the data used to calculate the equity premium and discussion of potential biases in the historical data, (2) analysis of the economic models, and (3) discussion of the implications of the findings for investors and for forecasts of the future equity premium.[1]

The equity risk premium determines asset allocations, projections of wealth, and the cost of capital, but we do not have a simple model that explains the premium.

Historical Returns on Stocks and Bonds

In this section, I present historical asset returns since 1802, define the equity premium, and discuss biases in the historical data that affect future estimates of the equity premium.

Jeremy J. Siegel is the Russell E. Palmer Professor of Finance at the Wharton School, University of Pennsylvania, Philadelphia.

Equity Returns. The historical returns on stocks, bonds, and bills and the equity risk premium for the U.S. markets from 1802 through 31 December 2004 are in **Table 1**.[2] Both the arithmetic mean of the annual data, which is the "expected return" used in the capital asset pricing model (CAPM), and the compound (or geometric) return, which is the return most often used by individual and professional investors, are given in Table 1.[3] The last columns display the equity risk premium in relation to both long-term U.S. government bonds and T-bills. Returns and premiums are broken down into two subperiods in Panel A, into three major subperiods in Panel B, and into the major bull and bear markets since World War II in Panel C.

The stability of the real (inflation-adjusted) return on stocks over all long periods is impressive.[4] The compound annual real return on equity has averaged 6.82 percent over the past 203 years and, as Panels B and C show, settled between 6.5 percent and 7.0 percent for each of the three major subperiods and for the post–World War II data. This return is about twice the growth of the economy and includes the risk premium above risk-free assets that investors have demanded to hold stocks.

When the period for which stock returns are analyzed shrinks to one or two decades, the real return on stocks can deviate substantially from the long-run average. Since World War II, returns in major market cycles have fluctuated from a 10.02 percent annual real equity return in the bull market of 1946–1965 to a –0.36 percent annual real equity return in the bear market of 1966–1981; in the great bull market of 1982–1999, the return doubled the 203-year average.

Fixed-Income Returns. The middle columns in Table 1 show that real bond returns, in contrast to stocks, have experienced a declining trend in the past two centuries. From 1802 through 2004, the average annual compound real return on long-term bonds was about half the equity return, but in the 19th century, real bond returns were nearly 5 percent. Since the end of World War II, the bond return has averaged less than 1.50 percent. The 3.31 percent average real return over the last two centuries is approximately equal to the real growth of the economy, but in the post–World War II period, real returns on bonds have fallen far below economic growth.[5]

The real return on short-dated T-bills has fallen even more sharply than the return on bonds over the past two centuries. For the entire period, real T-bill returns averaged 2.84 percent, 67 bps below the return on long-term bonds. Average short-term

Table 1. Historical Real Stock and Bond Returns and the Equity Premium

	Real Return						Stock Return minus Return on:			
	Stocks		Bonds		Bills		Bonds		Bills	
Period	Comp.	Arith.	Comp.	Arith.	Comp.	Arith.	Comp.	Arith.	Comp.	Arith.
A. Long periods to present										
1802–2004	6.82%	8.38%	3.51%	3.88%	2.84%	3.02%	3.31%	4.50%	3.98%	5.36%
1871–2004	6.71	8.43	2.85	3.24	1.68	1.79	3.86	5.18	5.03	6.64
B. Major subperiods										
1802–1870	7.02%	8.28%	4.78%	5.11%	5.12%	5.40%	2.24%	3.17%	1.90%	2.87%
1871–1925	6.62	7.92	3.73	3.93	3.16	3.27	2.89	3.99	3.46	4.65
1926–2004	6.78	8.78	2.25	2.77	0.69	0.75	4.53	6.01	6.09	8.02
C. Post–World War II full sample, bull markets, and bear markets										
1946–2004	6.83%	8.38%	1.44%	2.04%	0.56%	0.62%	5.39%	6.35%	6.27%	7.77%
1946–1965	10.02	11.39	–1.19	–0.95	–0.84	–0.75	11.21	12.34	10.86	12.14
1966–1981	–0.36	1.38	–4.17	–3.86	–0.15	–0.13	3.81	5.24	–0.21	1.51
1982–1999	13.62	14.30	8.40	9.28	2.91	2.92	5.22	5.03	10.71	11.38
1982–2004	9.47	10.64	8.01	8.74	2.31	2.33	1.46	1.90	7.16	8.32

Note: "Comp." stands for "compound"; "Arith." stands for "arithmetic."

rates were 34 bps above long-term rates for 1802–1870, but they were 57 bps below long rates from 1871 through 1925 and have been 156 bps below long rates since 1926.

The increase in the spread between long rates and short rates was caused partly by the increased liquidity of the T-bill market, which lowered short rates, and partly by the increase in the inflation premium investors have required on long-term bonds over much of the post–World War II period.

The Equity Premium. The decline in the real return on bonds, combined with the relative stability of the real return on equity, has increased the equity premium over time, as the last columns in Table 1 show. Over the 1802–2004 period, the equity risk premium as measured from compound annual returns and in relation to bonds rose (see Panel B) from 2.24 percent to 2.89 percent to 4.53 percent. Measured in relation to T-bills, the equity risk premium has increased even more.

The Risk-Free Rate: Long or Short? Should the equity risk premium be measured against the rate of short-term or long-term government bonds? In the simple representations of the CAPM, the risk-free rate is calculated against the rate on short-term risk-free assets, such as T-bills. When an intertemporal CAPM is used, however, a short rate may not be appropriate.[6] Investors should hedge against changes in investment opportunities, as represented by changes in the real risk-free rate. And in an intertemporal context, a risk-free asset can be considered an annuity that provides a constant real return over a long period of time.[7] The return on this annuity is best approximated by the returns on long-term inflation-indexed government bonds. In the United States, inflation-indexed government bonds were not introduced until 1997, so real returns on bonds before that date must be calculated *ex post* by subtracting inflation from nominal bond yields.

Calculation of the Equity Premium. The equity risk premium can be defined by the reference asset class, time period chosen, or method of calculating mean returns so as to take on a wide range of values. Its maximum value is calculated by using the *arithmetic* mean return of historical stock returns and subtracting the mean return on the highest-quality short-dated securities, such as T-bills. Measured in this way, the equity premium in the United States since 1802 has been 5.36 percent and since 1926, has been 8.02 percent. When *geometric* mean returns are used, the equity premium shrinks to 3.98 percent since 1802 and 6.09 percent since 1926. If we calculate the equity premium against long-dated (instead of short-term) bonds, the compound premium falls farther—to 3.31 percent over the past 202 years and 4.53 percent since 1926.

So, over the period from 1926 to the present, the premium can differ by 3.5 percentage points depending on whether long- or short-dated securities are used or arithmetic or geometric returns are calculated. Notwithstanding, the premium calculated by any of these methods far exceeds the magnitude derived in the Mehra–Prescott model.

Biases in Historical Equity Returns. In calculations of the equity risk premium, certain biases must be recognized: the international survivorship bias; failure to take transaction costs and diversification benefits into account; investor ignorance of risks, returns, and mean reversion; taxes and individuals' pension assets; and biases in the historical record of bond returns.

- *International survivorship bias.* Some economists claim that the historical real return on U.S. equities quite probably overstates the true expected return on stocks (Brown, Goetzmann, and Ross 1995). They maintain that the United States simply turned out to be the most successful capitalist country in history, a development that was by no means certain when investors were buying stock in the 19th and early 20th centuries.

Because the economic outcome in the United States was better than expected, U.S. returns may overstate the *expected* return on stocks. The cause is a phenomenon called "survivorship bias." This bias will exist whenever stock returns are recorded in successful equity markets, such as those in the United States, but omitted where stocks have faltered or disappeared outright, such as they did in Russia.

To address survivorship bias and to compile definitive series of long-term international stock returns, three U.K. economists—Dimson and Marsh from the London School of Business and Staunton from the U.K. statistical center—examined stock and bond returns over the past century in 16 countries. Their research, published in *Triumph of the Optimists: 101 Years of Global Investment Returns,* found that the superior returns on stocks over bonds is not characteristic of the U.S. market alone but exists in virtually all countries (see Dimson, Marsh, and Staunton 2002, 2004). **Figure 1** shows the average annual real stock, bond, and bill returns of the 16 countries they analyzed from 1900 through 2003.

Real equity returns ranged from a low of 1.9 percent in Belgium to a high of 7.5 percent in Sweden and Australia. Stock returns in the United States, although quite good, were not exceptional. U.S. stock returns were exceeded by the returns in Sweden, Australia, and South Africa.

If an equal investment had been placed in each of these markets in 1900, the average annual real return on stocks from 1900 through 2003 would have been 6.0 percent a year, not far below the U.S. return of 6.5 percent.[8] Furthermore, in the countries where real equity returns were low, such as Belgium, Italy, and Germany, real bond returns were also low, so the equity premium in Italy and Germany as measured against bonds was actually higher than the premium in the United States. In fact, the compound annual return of an equal amount invested in stocks in each country surpassed an identical amount in bonds in each country by 4 percent a year, only slightly less than the 4.6 percent equity risk premium found for the United States over the same time period.

When all the information was analyzed, the authors concluded:

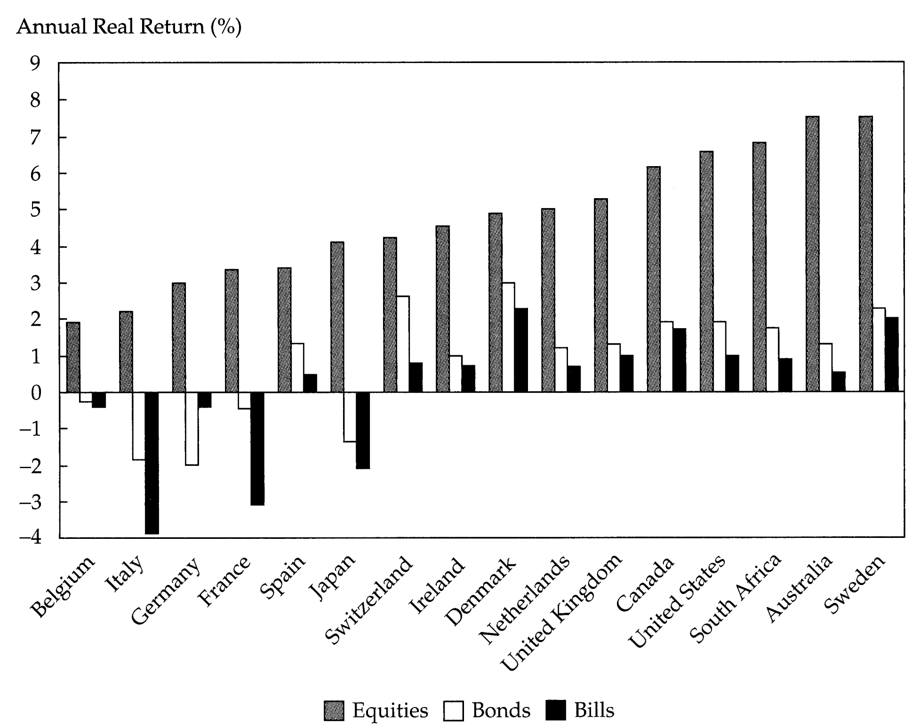
> While the U.S. and the U.K. have indeed performed well . . . there is no indication that they are hugely out of line with other countries. . . . Concerns about success and survivorship bias, while legitimate, may therefore have been somewhat overstated [and] investors may have not been materially misled by a focus on the U.S. (Dimson, Marsh, and Staunton 2002, p. 175)

The high historical equity premium is a worldwide, not just a U.S., phenomenon.[9]

Transaction costs and diversification. The returns used to calculate the equity premium are derived from published stock indices, but investors may not have realized these returns in their portfolios. Transaction costs in the equity markets were far higher over most of the period than they are today.

Low-cost indexed mutual and exchange-traded funds were not available to investors of the 19th century or most of the 20th century. Before 1975, brokerage commissions on buying and selling individual stocks were fixed by the NYSE at high levels. Moreover, it is not unreasonable to

Figure 1. Real Returns on International Assets, 1900–2003



assume that until recently, transaction costs involved with replicating a market portfolio with reinvested dividends subtracted 1–2 percentage points a year from stockholder returns.[10] So, the *realized* equity returns were probably much lower than those calculated from published data.

Investor ignorance of risks, returns, and mean reversion. Because data on long-term stock returns were not available until the second half of the 20th century, investors in the past were probably ignorant of the true risks and returns from holding stocks and may have underestimated the return and/or overestimated the risk of equities. When Fisher and Lorie (1964) first documented long-term returns in the 1960s, many economists were surprised that even when the Great Depression was included, stocks yielded such a high rate of return.

Another advantage of stocks that until recently was not recognized is the evidence of *mean reversion* of long-term equity returns.[11] In the early development of capital asset pricing theory, financial returns were modeled as random walks whose risk increased as the square root of the time period. But examination of long-term data strongly suggests a predictable component of stock returns that makes the returns less variable over long periods than they would be if mean reversion did not exist. Mean reversion increases the desirability of stocks as assets for long-term investors.

Ignorance of the historical risks and returns of various asset classes may have led to a general underpricing of equities as an asset class. This result, in turn, may have raised realized returns higher than would be justified if stocks were priced by investors with full knowledge of the distribution of stock returns.[12]

Pension assets and taxes. The evolution of U.S. federal tax policy also may have influenced stock returns. The tremendous increase in tax-sheltered plans over the past several decades has greatly increased the demand for equities. For example, in 1974, ERISA established minimum standards for pension plans in private industry and allowed equities to play a greatly expanded role in asset accumulation.

McGrattan and Prescott (2003) argued that the increase in tax-sheltered savings has led to a significant drop in the average tax rate on equities. This drop may have boosted stock returns and, to the extent that stocks substituted for bonds, lowered the real return on fixed-income assets.

Biases in historical bond returns. Real government bond returns may have been biased downward in the period since 1926, especially since World War II. Bondholders clearly did not anticipate the double-digit inflation of the 1970s and 1980s.

Table 1 shows the extraordinarily poor bond returns in the 35 years following World War II. Of course, when inflation was brought down in the 1980s and 1990s, interest rates returned to the levels of the immediate postwar period. But the resulting bull market in bonds did not offset the losses of the inflationary 1960s and 1970s because, although the inflation rate returned to its earlier level, the *price level* did not. So, over the entire inflation cycle, bondholders suffered a permanent loss of return. This phenomenon is one reason real bond returns since World War II have averaged only 1.4 percent, less than half their historical level.[13]

Models of the Equity Premium

The biases just discussed have probably raised the historical return on equities and, therefore, the historical value of the equity risk premium. Nevertheless, accounting for these biases is unlikely to reduce the premium to the level that Mehra and Prescott maintain is consistent with reasonable levels of risk aversion. So, we are compelled to analyze whether the assumptions of the models used to describe investor behavior are, in fact, reasonable representations of investor and financial market behavior.

The equity premium puzzle is centered on the "reasonable" level of risk aversion for investors. Recall that risk premiums exist because individuals are assumed to have declining marginal utility of consumption. How fast this utility declines measures the investor's degree of risk aversion. In early risk models, the investor's utility function, U, was assumed to be a function of wealth, W, such that

$$U(W) = \left[\frac{1}{(1-A)}\right] W^{(1-A)}. \quad (1)$$

The parameter A is the coefficient of relative risk aversion, or the percentage change (elasticity) of the marginal utility of wealth caused by a 1 percent change in the level of wealth. In other words, A is directly related to the pain felt by investors when their wealth falls.

With this utility function, and under the assumption that returns are lognormally distributed, the arithmetic equity premium, EP, can be approximated by

$$EP \approx A(\sigma^2), \quad (2)$$

where σ is the standard deviation of returns on an investor's portfolio. If we use 0.18 as the standard deviation of annual stock market returns and an (arithmetic) equity risk premium of 8 percent as measured from annual data since 1926, we obtain a level of risk aversion, A, of 2 or 3.[14]

These levels of risk aversion produced by the early models seemed reasonable. With a risk aversion of 2, an individual would be willing to pay 4 percent of his wealth to insure against an equal probability of a 20 percent rise or 20 percent fall in wealth. If A equals 3, this insurance payment would be 5.6 percent of wealth.

But Equation 1 is not correctly specified. Economists knew that wealth is a proxy for consumption, which is the correct variable to put into the utility function. Putting consumption into the utility function led to the development of the "consumption CAPM" (CCAPM) popularized by Breeden (1979).

There is an important empirical difference between the consumption-based CAPM and the wealth-based CAPM. Per capita consumption, as measured by national income account statistics, fluctuates far less than the value of wealth. The standard deviation of the growth of consumption is only about 4 percent, so the variance of changes in the stock market is almost 20 times greater than the variance of the changes in consumption.

If we plug the variance of consumption of 0.16 percent and an equity premium of 8 percent into Equation 2, we find a risk aversion of 50. If investors were really this risk averse, they would pay an insurance premium of 17 percent to avoid an equal probability of a 20 percent rise or fall in their wealth. For investors to act this risk averse is implausible. In other words, if individuals actually have a risk aversion coefficient of 2 or 3, the equity risk premium implied in the CCAPM is much smaller, on the order of 0.3–0.4 percent. The intuition here is that historical changes in consumption are not large enough to significantly alter utility, so investors are willing to take nearly a "fair bet" with stocks.[15]

Another way of looking at this issue is that the standard CAPM assumes that changes in wealth cause equal changes in consumption, but in reality, movements in the stock market are not associated with dramatic changes in consumption. Any risk that is not strongly correlated with consumption should not require a large risk premium, and empirically, the returns on equities fall into that category.[16]

The equity premium puzzle was not the only anomaly implied by the consumption CAPM. Weil (1989) showed that not only did the CCAPM imply that the historical equity premium was too large, but it also implied that the historical real rate of return on bonds, given economic growth and reasonable risk-aversion parameters, was far too small. This anomaly was called the "risk-free rate puzzle." These two puzzles were related to the "excess volatility puzzle," which had been explored earlier by Shiller (1981), who showed that stock prices have been too volatile to be explained by changes in subsequent dividends.

These puzzles are caused by the fact that the stock market has fluctuated far more than the underlying economic variables, such as aggregate consumption or GDP.

Finding the Model That Fits the Data

Before attempting to change the basic model summarized by Equation 1 with consumption substituting for wealth, I should note that some economists believe that the high levels of risk aversion implied by the model are not necessarily unreasonable. Kandel and Stambaugh (1991) pointed out that, although high levels of risk aversion may lead to unreasonable behavior with respect to large changes in consumption, the behavior may not be implausible for small changes in wealth. For example, to avoid a 50/50 chance of your consumption rising or falling by 1 percent if your coefficient of risk aversion is 10, you would pay 5 percent of the gamble. Even if risk-aversion coefficient A is as high as 29, which best fits the data in the Kandel–Stambaugh model, an investor would pay only 14.3 percent of the gamble to avoid the risk of a 1 percent rise or fall in wealth. Neither of these actions appears unreasonable.

Fama, agreeing that a large risk-aversion coefficient is not necessarily a puzzle, stated that

> a large equity premium says that consumers are extremely averse to small negative consumption shocks. This is in line with the perception that consumers live in morbid fear of recessions (and economists devote enormous energy to studying them) even though, at least in the post war period, recessions are associated with small changes in per capita consumption. (1991, p. 1596)

In evaluating these arguments, however, remember that in the domain of retirement savings, the stakes are large relative to wealth or yearly consumption. A typical faculty member at age 55 saving, say, 10 percent of her salary a year might well have half or more of her wealth (including future earnings) in her retirement account. Similarly, university endowments are a substantial portion of the wealth of private universities. And even with mean reversion of equity returns, the 10-year to 20-year standard deviation of equity returns is substantial. So, we seem to be back in the high-stakes category, where high values of risk aversion lead to absurd behavior.

Changes in the Utility Function. In an attempt to solve the puzzle, most economists have been driven to modify the consumption-based utility function represented by Equation 1 to justify a higher equity premium without requiring an implausibly high level of risk aversion. A popular generalization of Equation 1, pioneered by Epstein and Zin (1989), breaks the rigid link between risk aversion (investor reaction to changes in consumption over a *given period of time*) and the reaction to changes in consumption over time, called the *intertemporal rate of substitution*, which affects the real rate of interest. This class of utility functions has been fruitful in explaining low real rates but does not go far in explaining the equity premium.

Another line of research makes utility a function not only of current consumption but also of some "benchmark" level of consumption. If the benchmark is taken to be prior levels of consumption, then individuals are taken to be sensitive not only to their level of consumption today but also to how it has changed from yesterday. Thus, individuals are assumed to take time to adjust to new levels of consumption, a behavior that can be described as "habit formation."

Constantinides (1990) showed that habit formation makes an investor more risk averse to a short-run change in consumption, leading to higher "short-run" risk aversion than "long-run" risk aversion. Evidently, once one has tasted the good life, it is difficult to adjust one's consumption downward. A similar approach was taken by Campbell and Cochrane (1999), who claimed that utility is a function of consumption over and above some habit that is slow to change. Therefore, in a recession, risk aversion increases markedly even though in absolute terms, recessions exhibit relatively small declines in consumption. The equity premium, as well as all other risk premiums, does indeed increase in recessionary periods.

Abel (1990) examined asset pricing when an individual's utility is derived not only from the individual's own consumption but also *relative* to the consumption of others around them—what he termed "catching up with the Joneses." This utility function is less risk averse if everyone's income moves up and down together, but when individuals compare their living standards with others', the comparison makes individuals act very risk averse. This utility function helps solve the real rate puzzle but is not much help in explaining the equity premium.[17]

An alternative approach, elaborated by Benartzi and Thaler (1995), is built on the "cumulative prospect theory" proposed by Tversky and Kahneman (1992). Prospect theory shares the claim that utility is based on benchmarks, so today's level of consumption is important, but prospect theory, which is a pioneering model in behavioral finance, asserts that asset *returns*, rather than consumption or wealth, are arguments of the utility function. In these models, investors dislike losses much more intensely than they like gains. When the utility function is based on *changes* in wealth rather than *levels* of wealth, investors are referred to as "loss averse" rather than "risk averse."[18]

When investors have these loss-averse preferences, their attitudes toward risky assets depend crucially on the time horizon over which returns are evaluated. For example, loss-averse investors who compute the values of their portfolios every day would find investing in stocks unattractive because stock prices fall almost as often as they rise. Investors who check returns less frequently have a higher probability of seeing positive returns. The concept of loss-averse preferences explains why individuals are so risk averse in the short run, what Benartzi and Thaler called "myopic loss aversion."

Uncertain Labor Income. The previous models assumed that the only important source of uncertainty is the return on equity. A more realistic way to model uncertainty would be to recognize that labor income is also uncertain. This fact can markedly change investors' behavior toward the risks in financial markets.

Uncertain labor income may explain why risk aversion increases in a recession; it is well known that unemployment and the number of layoffs affect workers' decisions. During recessions, stocks frequently sell at large discounts relative to their long-term values, a factor that increases long-run equity returns.

The inability to borrow large sums against labor income also means that many workers, especially young workers, are not able to hold as much equity as they would like, even though their "human capital," measured as the value of their future labor income, is high. Constantinides, Donaldson, and Mehra (2002) reported that this phenomenon can have important consequences for asset pricing. Older workers do hold equity, but this age cohort displays greater risk aversion than younger workers because older workers have much more limited ability to offset portfolio losses by changing their work effort. As a result, the economy in general displays the greater risk aversion of the older generation, for whom future consumption is more geared to the level of financial assets than to income. Indeed, Mankiw and Zeldes (1991) found that large stockholders' consumption reflects a larger sensitivity to market fluctuations than does the consumption of smaller stockholders.

Modeling the Risks to Consumption and Equities. Another path to justifying the equity risk premium, rather than changing the form of the utility function, is to reexamine the statistical properties of consumption and stock returns. The standard approach is to assume that both the growth of consumption and the return on stocks are stochastic processes marked by lognormal distributions with constant expected returns. Although this specification is analytically tractable and reasonably replicates the behavior of the historical data, it may not be correct.

Weitzman (2004) argues in a working paper that we do not know the exact distributions of output in the economy, so treating the historically estimated means and standard deviations as known parameters is incorrect. Uncertainty about the true means and variances of the distribution signifies that the probability distributions of consumption and stock returns have fatter tails than assumed in the lognormal distribution.

We know that stock returns do, in fact, have far fatter tails than implied by lognormality. If lognormality prevailed, the probability of the 19 percent decline in the S&P 500 Index that occurred on 19 October 1987 would be less than 1 in 10^{71}, so even if we had had billions of exchanges operating daily for the last 12 billion years (the estimated age of the universe), there would be virtually no chance of observing this event. Yet, the decline did occur, and it may have dramatically increased investors' perceptions of equity risk.

Weitzman shows that, in the absence of risk-free assets, these fatter-tailed distributions alter the analytics of the equity premium dramatically. Instead of yielding an extremely low equity premium, these distributions yield an arbitrarily high equity premium for any level of risk aversion. Furthermore, this model has the ability to explain a low risk-free rate *and* the "excess volatility" of the stock market.

This research is not unrelated to the earlier studies of Rietz (1988), who speculated shortly after Mehra and Prescott's research that investors fear a lurking "disaster state" of extreme negative consumption that has not yet been realized. Such fear would lead to a higher equity premium.[19] Recently, Barro (2005) found strong support for this theory in the data for international markets.

In a similar vein, Bansal and Yaron (2004) rewrote the stochastic properties of the consumption and dividend growth models. Instead of modeling consumption growth as uncorrelated through time, they assumed it has a small long-run predictable component that is affected by past growth. So, a shock to consumption influences its expected growth as well as the expected growth of dividends many years into the future, which can have a dramatic impact on the valuation of equities.[20] When this consumption process is combined with time-varying variance, the Bansal–Yaron model, like Weitzman's approach, has the capability of explaining all the asset pricing puzzles.[21]

Practical Applications

The practitioner might ask: How does the equity premium puzzle matter to investors? This question should be analyzed in the following way.

If the equity premium should be only a fraction of 1 percent, as the basic economic model suggests, then either stocks should be priced much higher or bonds should be priced much lower than they have been on a historical basis.[22] If stock prices rose and bond prices fell, the result would lower the forward-looking returns on equities and raise returns on fixed-income assets, thereby lowering the equity premium. Clearly, if investors believe this narrower premium will prevail at some time in the future, they should be fully invested in stocks now.

But this scenario is highly unlikely to occur. Although the future equity premium is likely to be somewhat lower than in the past, few believe investors will hold stocks if their expected return is only a fraction of a percent above the return of risk-free assets.

Yet, we should not dismiss the equity premium puzzle. The search for the right model has yielded insights that can give practitioners guidance in structuring their clients' portfolios. One promising area is the work on habit formation, which implies that there may be a significant difference in an investor's short-term and long-term attitudes toward risk. This research suggests that an advisor may find it worthwhile to explore the investor's reaction to lowering consumption in a short time frame versus lowering it in a longer time frame, when other adjustments can be made to ease the impact of a reduced standard of living.

A related issue is the importance of examining labor income as a component of portfolio choice. Individuals whose labor income is uncertain and whose borrowing capabilities are low should hold a lower allocation of equities. Those with highly marketable skills should hold a higher fraction in equities. Those who are near retirement and have no flexibility to change their labor income will be more risk averse than investors with marketable labor skills.

A high equity premium can arise from assuming that investors demand a minimum level of consumption that must be attained in any investment plan, no matter what the time period to adjust. The effect is equivalent to assuming that risk aversion becomes extremely high at low levels of consumption. This approach has given rise to the growth of "liability investing," in which investors, especially those approaching retirement, fund what they deem absolute minimum expenditures with risk-free assets, such as Treasury Inflation-Indexed Securities (informally called TIPS), with the remainder being subject to the usual risk and return trade-offs (see Waring 2004).

Investors who suffer from myopic loss aversion, the condition in which the downs in the market deliver much more pain than the ups deliver pleasure, should be advised to set their best allocations and then assess the value of their portfolios infrequently. Blind trusts controlled by outside advisors might be the best strategy for the investors who are particularly sensitive to losses.

Financial planners must also evaluate their clients' fears of remote but catastrophic events and evaluate the likelihood of such events. In some economic states, such as a terrorist strike or a nuclear attack, equities could suffer extreme losses. Practitioners should note that these events will also affect the value of government bonds, so what are considered risk-free assets may even no longer exist.[23] War and other conflicts that destroy wealth also cannot be ruled out. Furthermore, over a very long horizon, there is the possibility that capitalism as a form of economic organization may cease to exist and that the wealth of the propertied classes will be expropriated. For investors with fears of these remote, yet not inconceivable, events, a financial advisor must determine whether the equity premium is sufficient to overcome the outcomes.

Future of the Equity Risk Premium

Despite the fact that the models that economists taught in their classes predicted a small equity premium, most academic economists, even at the peak of the bull market in 2000, maintained a personal estimate of the equity premium (which, presumably, they taught to students) close to the historical mean realized premium since 1926—that is, about 6 percent (compound) or 8 percent (arithmetic) over T-bills.

For his 2000 paper, Welch surveyed a large number of academic economists, who estimated the arithmetic premium of stocks over short-term bonds at 7 percent, about 100 bps below the 1926–2004 average.[24] If we subtract 2 percentage points to convert to the geometric average and then subtract a further 150 bps to convert from short-run to long-run bonds, we obtain a geometric equity premium of stocks over bonds of about 3.5 percent.

Professional money managers apparently have a lower estimate of the equity risk premium than do academics. At a CFA Institute conference I spoke to in early 2004, Peter Bernstein—noted author, money manager, and an organizer of the conference—asked the large crowd of professional investors whether they would be inclined to hold in their portfolios a preponderance of equity over fixed income if they knew that the equity premium was 3 percent. A majority raised their hands. When he asked the same question with a 2 percent premium, most of the audience did not.[25]

I noted in the opening of this article that persuasive reasons support a lower forward-looking real return on equity than the return found in the historical data. The sharp drop in the cost of acquiring and maintaining a diversified portfolio of common stocks, not only in the United States but now worldwide, should increase the price of equities and lower their future return. If we assume these annual costs have been brought down by 100 bps, then the future real return on equities should be 5.5–6.0 percent, about 1 percentage point lower than the historical range of 6.5–7.0 percent. Although these returns are below the historical average calculated from indices, investors today will receive the same realized return from stocks as they obtained earlier when trading costs were higher.

For bonds, the question is whether real future returns should be higher than the 2.25 percent average recorded since 1926. Until recently, I believed that the answer was unambiguously yes. The historical real return on bonds was biased downward by the inflation of the 1970s. Indeed, when TIPS were issued in 1997, their real yield was 3.5 percent, and it climbed to more than 4 percent in 2000. If we assume future real bond returns will be 3.5 percent and real stock returns will be between 5.5 percent and 6 percent, the equity premium will be between 2 percent and 3 percent, a level that would leave most money managers satisfied with their equity allocations.

But in the last few years, the real return on protected government bonds has dropped sharply. TIPS yields, which had been as high as 3 percent in the summer of 2002, fell to 1.5 percent in 2005. The causes of the drop are not well understood but may be related to such factors as fear of a decline in growth because of the decline in the number of workers, the increased risk aversion of an aging population, the excess of saving over investment, manifesting itself through the demand for U.S. government bonds from developing Asian countries, or the increased demand for fixed-income assets by pension funds seeking to offset their pension liabilities. Another possibility is that bondholders believe central banks will keep inflation low, so they view government bonds as true hedges against disaster scenarios ranging from armed conflict to terrorist attacks—and even natural disasters.

If the equity premium is 2–3 percent and real bond yields remain at 1.5 percent, the projected real return on stocks is only about 4 percent. Some noted analysts believe that real stock returns will indeed be this low because this return comports with a 2 percent dividend yield plus the 2 percent long-term real growth of per share dividends found in long-run stock data (Bernstein and Arnott 2003).

I believe, however, that this forecast of real stock returns is too low. First, future dividend growth should be higher than the historical average because the dividend payout ratio has fallen dramatically, which enables companies to use retained earnings to finance growth.[26] Second, future real stock returns can be predicted by taking the earnings yield, which is the inverse of the well-known P/E. This approach works extremely well with long-run data because the average historical P/E of 15 has corresponded to a 6.7 percent real return on stocks. The P/E taken from data in August 2005 points to a 5.5–6.0 percent real stock return. As mentioned earlier, the higher level of stock prices relative to earnings is justified by the steep decline in the costs of holding a fully diversified equity portfolio.

Finally, I believe that the pessimism about future economic growth is unwarranted. In my opinion, the negative impact of the aging of the developed world's population will be more than offset by accelerating growth in the developing world, which will lead to rapid worldwide growth over the next several decades.[27] Forward-looking equity returns of an internationally diversified portfolio should therefore be in the range of 5.5–6.0 percent. If the real return on bonds remains in the 1.5–2.0 percent range, because of increased risk aversion or other factors unrelated to economic growth, then the equity risk premium has probably risen to a level that comports with the post-1926 data.

Conclusion

The equity premium is a critical number in financial economics. It determines asset allocations, projections of retirement and endowment wealth, and the cost of capital to companies. Economists are still searching for a simple model that can justify the premium in the face of the much lower volatility of aggregate economic data. Although there are good reasons why the future equity risk premium should be lower than it has been historically, projected compound equity returns of 2–3 percent over bonds will still give ample reward for investors willing to tolerate the short-term risks of stocks.

Notes

1. Many excellent academic reviews of the equity premium puzzle are available. Cochrane (2005) of the University of Chicago has provided a complete updated review.
2. The stock series is from a combination of sources. Data for 1802–1871 are from Schwert (1990); data for 1871–1925 are from Cowles (1938); data for 1926–2004 are from the CRSP capitalization-weighted indexes of all NYSE, Amex, and NASDAQ stocks. More extensive descriptions of the data can be found in Siegel (2002).
3. As an approximation, the geometric return is equal to the arithmetic return minus one-half the variance of the return. For a fuller description, see the subsection "Calculation of the Equity Premium."
4. Smithers and Wright (2000) called this stable long-term return "Siegel's Constant."
5. Theoretically, real interest rates do not necessarily equal growth. The real rate is also a function of the time rate of discount and the level of risk aversion.
6. See Merton (1973) for a description of the intertemporal CAPM.
7. Campbell and Viceira (2002) indicated that the yield on the 10-year U.S. inflation-linked bond would be the closest in duration to the indexed annuity, especially for someone approaching retirement.
8. Mathematically, the average return of an equally weighted world portfolio is higher than the average equity return in each country.
9. In fact, *Triumph of the Optimists* may have actually *understated* long-term international stock returns. The U.S. stock markets and other world markets for which we have data did very well in the 30 years prior to 1900, which is when their study began. U.S. returns measured from 1871 outperformed returns taken from 1900 by 32 bps. Data from the United Kingdom show a similar pattern.
10. Before commissions were deregulated in May 1975, a typical trade—say, 100 shares at $30—paid a commission of $58.21, almost 2 percent of market value. Small odd-lot trades resulting from reinvesting dividends could cost, considering odd-lot premiums, as much as 4 percent.
11. See Poterba and Summers (1988) for early research on mean reversion and Cochrane (1999) for evidence of stock return predictability.
12. Abel (2002) explored the implications for the equity risk premium when investors had incorrect information on the distributions of returns.
13. Recently, real bond returns have fallen sharply, which is discussed later.
14. See Friend and Blume (1975) for an earlier derivation of the risk-aversion parameter.
15. Arrow (1965) showed that for small risks, investors should be risk neutral, requiring little or no premium.
16. When consumption and stock returns are not perfectly correlated, $EP = \sigma_c \sigma_W \rho_{c,W}$, where σ_c is the standard deviation of consumption, σ_W is the standard deviation of stocks, and $\rho_{c,W}$ is the correlation coefficient between the two. Because empirically ρ is about 0.2, this equation leads to approximately the same estimate of risk aversion as does the CCAPM (see Cochrane 2005).
17. Once Abel (1999) added leverage, the equity premium was better estimated.
18. In the standard model, loss aversion is equivalent to a "kink" in the utility function at the current level of consumption. The loss in utility when consumption drops below the kink is greater than the gain when consumption is above, even for tiny changes in consumption.
19. Mehra and Prescott (1988), criticizing Rietz's research, noted that a disaster state was very likely to be realized in the more than 100 years of data that Mehra and Prescott analyzed.
20. The intuition here comes from the Gordon model of stock price determination, in which small changes in the growth rate of dividends have a large impact on stock prices.
21. Note that in reconciling the volatility of stocks with underlying macroeconomic variables, the compilation of national income accounts requires a large amount of estimation and smoothing of past data, and averaged data on any index lower its volatility. As for estimation, it is well known that the "appraised" value of real estate is far more stable than the value of securities that *represent* similar assets, such as REITs.
22. Indeed, a best-selling book by James Glassman and Kevin Hassett (1999) on the stock market, *Dow 36,000*, marketed at the peak of the last bull market, maintained this thesis and predicted that stocks would have to increase fourfold to bring their real yields down to those of bonds.
23. Perhaps this fear explains why gold continues to be popular despite the fact that in portfolio models, precious metals are often dominated by stocks and inflation-protected bonds.
24. These academics predicted that other academics' estimates were higher—in the 7.5–8.0 percent range.
25. The conference was "Points of Inflection: Investment Management Tomorrow"; a webcast of the Bernstein presentation is available at www.cfawebcasts.org. Rob Arnott has been doing such surveys for a number of years and has communicated to me that most of the institutional money managers would be satisfied with an equity premium measured against bond returns of 2–3 percent (see Arnott and Bernstein 2002).
26. If retained earnings can be invested at the same rate of return as required by equity investors, a drop in the dividend yield will produce an equal rise in the future growth of dividends (see Siegel 2002). Arnott and Asness (2003), believing that company managers squander retained earnings on low-return projects, rejected my contention that real dividends will grow faster in the future.
27. See Siegel (2005) for support for these statements.

References

Abel, Andrew B. 1990. "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review*, vol. 80, no. 2 (May):38–42.

———. 1999. "Risk Premia and Term Premia in General Equilibrium." *Journal of Monetary Economics*, vol. 43, no. 1 (February):3–33.

———. 2002. "An Exploration of the Effects of Pessimism and Doubt on Asset Returns." *Journal of Economic Dynamics and Control*, vol. 26, nos. 7–8 (July):1075–92.

Arnott, Robert D., and Clifford S. Asness. 2003. "Surprise! Higher Dividends = Higher Earnings Growth." *Financial Analysts Journal*, vol. 59, no. 1 (January/February):70–87.

Arnott, Robert D., and Peter L. Bernstein. 2002. "What Risk Premium Is 'Normal'?" *Financial Analysts Journal*, vol. 58, no. 2 (March/April):64–85.

Arrow, Kenneth A. 1965. *Aspects of the Theory of Risk-Bearing*. Helsinki, Finland: Yrjö Hahnsson Foundation.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance*, vol. 59, no. 4 (August):1481–1509.

Barro, Robert J. 2005. "Rare Events and the Equity Premium." National Bureau of Economic Research Working Paper No. 11310 (May).

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, vol. 110, no. 1 (February):73–92.

Bernstein, William J., and Robert D. Arnott. 2003. "Earnings Growth: The Two Percent Dilution." *Financial Analysts Journal*, vol. 59, no. 5 (September/October):47–55.

Breeden, Doug. 1979. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics*, vol. 7, no. 3 (September):265–296.

Brown, S.J., W.N. Goetzmann, and S.A. Ross. 1995. "Survival." *Journal of Finance*, vol. 50, no. 3 (July):853–873.

Campbell, John Y., and J.H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy*, vol. 107, no. 2 (April): 205–251.

Campbell, John Y., and Luis M. Viceira. 2002. *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. New York: Oxford University Press.

Cochrane, John. 1999. "New Facts in Finance." *Economic Perspectives* (Federal Reserve Bank of Chicago), vol. 23, no. 3 (3rd Quarter):36–58.

———. 2005. "Financial Markets and the Real Economy." NBER Working Paper No. 11193 (March).

Constantinides, George M. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, vol. 98, no. 3 (June):519–543.

Constantinides, George M., J.B. Donaldson, and R. Mehra. 2002. "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle." *Quarterly Journal of Economics*, vol. 117, no. 1 (February):269–296.

Cowles, A. 1938. *Common-Stock Indexes, 1871–1937*. Bloomington, IN: Principia Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton. 2002. *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

———. 2004. *Global Investment Returns Yearbook 2004*. ABN-AMRO (February).

Epstein, L.G., and S.E. Zin. 1989. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica*, vol. 57, no. 4 (July):937–969.

———. 1991. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis." *Journal of Political Economy*, vol. 99, no. 2 (April): 263–286.

Fama, Eugene. 1991. "Efficient Capital Markets: II." *Journal of Finance*, vol. 46, no. 5 (December):1575–1617.

Fisher, Lawrence, and James H. Lorie. 1964. "Rates of Return on Investment in Common Stocks." *Journal of Business*, vol. 37 (January):1–21.

Friend, Irwin, and Marshall Blume. 1975. "The Demand for Risky Assets." *American Economic Review*, vol. 65, no. 5 (December):900–922.

Glassman, James, and Kevin Hassett. 1999. *Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market*. New York: Crown Business.

Kandel, Shmuel, and Robert F. Stambaugh. 1991. "Asset Returns and Intertemporal Preferences." *Journal of Monetary Economics*, vol. 27, no. 1 (February):39–71.

MaCurdy, Thomas, and John Shoven. 1992. "Accumulating Pension Wealth with Stocks and Bonds." Working paper, Stanford University (January).

Mankiw, N. Gregory, and Stephen P. Zeldes. 1991. "The Consumption of Stockholders and Non-Stockholders." *Journal of Financial Economics*, vol. 29, no. 1 (March):97–112.

McGrattan, Ellen R., and Edward C. Prescott. 2003. "Average Debt and Equity Returns: Puzzling?" Staff Report 313, Federal Reserve Bank of Minneapolis.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–162.

———. 1988. "The Equity Risk Premium: A Solution?" *Journal of Monetary Economics*, vol. 22, no. 1 (July):133–136.

Merton, Robert C. 1973. "An Intertemporal Capital Asset Pricing Model." *Econometrica*, vol. 41, no. 4 (October):867–887.

Poterba, James, and Lawrence H. Summers. 1988. "Mean Reversion in Stock Returns: Evidence and Implications." *Journal of Financial Economics*, vol. 22, no. 1 (April):27–60.

Rietz, T.A. 1988. "The Equity Risk Premium: A Solution." *Journal of Monetary Economics*, vol. 22, no. 1 (July):117–131.

Schwert, William. 1990. "Indexes of United States Stock Prices from 1802 to 1987." *Journal of Business*, vol. 63, no. 3 (July):399–426.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *American Economic Review*, vol. 71, no. 3 (June):421–436.

Siegel, Jeremy J. 2002. *Stocks for the Long Run*. 3rd ed. New York: McGraw-Hill.

———. 2005. *The Future for Investors*. New York: Crown Publishing.

Siegel, Jeremy J., and Richard H. Thaler. 1997. "Anomalies: The Equity Premium Puzzle." *Journal of Economic Perspectives*, vol. 11, no. 1 (Winter):191–200.

Smithers, Andrew, and Stephen Wright. 2000. *Valuing Wall Street*. New York: McGraw-Hill.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, vol. 5, no. 4 (October):297–323.

Waring, M. Barton. 2004. "Liability-Relative Investing." *Journal of Portfolio Management*, vol. 30, no. 4 (Summer):8–19.

Weil, Philippe. 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics*, vol. 24, no. 3 (November):401–421.

Weitzman, Martin L. 2004. "The Bayesian Equity Premium." Working paper, Harvard University (August).

Welch, Ivo. 2000. "Views of Financial Economists on the Equity Premium and on Professional Controversies." *Journal of Business*, vol. 73, no. 4 (October):501–538.