
The Equity Premium: Why Is It a Puzzle?

(as corrected February 2003)

Rajnish Mehra

This article takes a critical look at the equity premium puzzle—the inability of standard intertemporal economic models to rationalize the statistics that have characterized U.S. financial markets over the past century. A summary of historical returns for the United States and other industrialized countries and an overview of the economic construct itself are provided. The intuition behind the discrepancy between model prediction and empirical data is explained. After detailing the research efforts to enhance the model’s ability to replicate the empirical data, I argue that the proposed resolutions fail along crucial dimensions.

Almost two decades ago, Edward Prescott and I (see Mehra and Prescott 1985) challenged the profession with a poser: The historical U.S. equity premium (the return earned by a risky security in excess of that earned by a relatively risk free U.S. T-bill) is an order of magnitude greater than can be rationalized in the context of the standard neoclassical paradigm of financial economics. This regularity, dubbed “the equity premium puzzle,” has spawned a plethora of research efforts to explain it away. In this article, I take a retrospective look at the puzzle and critically evaluate the various attempts to solve it.¹

Empirical Facts

Historical data provide a wealth of evidence documenting that for more than a century, U.S. stock returns have been considerably higher than returns for T-bills. As **Table 1** shows, the average annual real return (that is, the inflation-adjusted return) on the U.S. stock market for the past 110 years has been about 7.9 percent. In the same period, the real return on a relatively riskless security was a paltry 1.0 percent. The difference between these two returns, 6.9 percentage points (pps), is the equity premium. This statistical difference has been even more pronounced in the post-World War II period. Siegel’s (1998) data on U.S. stock and bond returns going back to 1802 reveal a similar, although somewhat smaller, premium for the past 200 years.

Rajnish Mehra is professor of finance at the University of California, Santa Barbara, a research associate of the National Bureau of Economic Research, and senior investment advisor to Vega Asset Management, New York.

Furthermore, this pattern of excess returns to equity holdings is not unique to U.S. capital markets. **Table 2** confirms that equity returns in other developed countries also exhibit this historical regularity when compared with the return to debt holdings. The annual return on the U.K. stock market, for example, was 5.7 percent in the post-WWII period, an impressive 4.6 pp premium over the average bond return of 1.1 percent. Similar statistical differences have been documented for France, Germany, and Japan. And together, the United States, the United Kingdom, Japan, Germany, and France account for more than 85 percent of capitalized global equity value.

The dramatic investment implications of the differential rates of return can be seen in **Table 3**, which maps the enormous disparity in capital appreciation of \$1 invested in different assets for 1802–1997 and for 1926–2000.²

This kind of long-term perspective underscores the remarkable wealth-building potential of the equity premium and explains why the equity premium is of central importance in portfolio allocation decisions, in making estimates of the cost of capital, and in the current debate about the advantages of investing Social Security funds in the stock market.

A Premium for Bearing Risk?

Why has the rate of return on stocks been significantly higher than the rate of return on relatively risk free assets? One intuitive answer is that stocks are “riskier” than bonds and investors require a premium for bearing this additional risk. Indeed, the standard deviation of the returns to stocks (about 20 percent a year historically) is larger than that of the returns to T-bills (about 4 percent a year),

Table 1. U.S. Returns, 1802–2000

Period	Mean Real Return		
	Market Index	Relatively Riskless Security	Risk Premium
1802–1998	7.0%	2.9%	4.1 pps
1889–2000	7.9	1.0	6.9
1926–2000	8.7	0.7	8.0
1947–2000	8.4	0.6	7.8

Sources: Data for 1802–1998 are from Siegel (1998); for 1889–2000, from Mehra and Prescott (1985), updated by the author. The rest are the author's estimates.

Table 2. Returns for Selected Countries, 1947–98

Country	Period	Mean Real Return		
		Market Index	Relatively Riskless Security	Risk Premium
United Kingdom	1947–99	5.7%	1.1%	4.6 pps
Japan	1970–99	4.7	1.4	3.3
Germany	1978–97	9.8	3.2	6.6
France	1973–98	9.0	2.7	6.3

Sources: Data for the United Kingdom are from Siegel; the rest of the data are from Campbell (forthcoming 2003).

Table 3. Terminal Value of \$1 Invested

Investment Period	Stocks		T-Bills	
	Real	Nominal	Real	Nominal
1802–1997	\$558,945	\$7,470,000	\$276	\$3,679
1926–2000	266.47	2,586.52	1.71	16.56

Sources: Ibbotson Associates (2001); Siegel (1998).

so obviously, stocks are considerably riskier than bills.

But are they? **Figure 1** illustrates the variability in the annual real rate of return on the S&P 500 Index (Panel A) and a relatively risk free security (Panel B) over the 1889–2000 period.

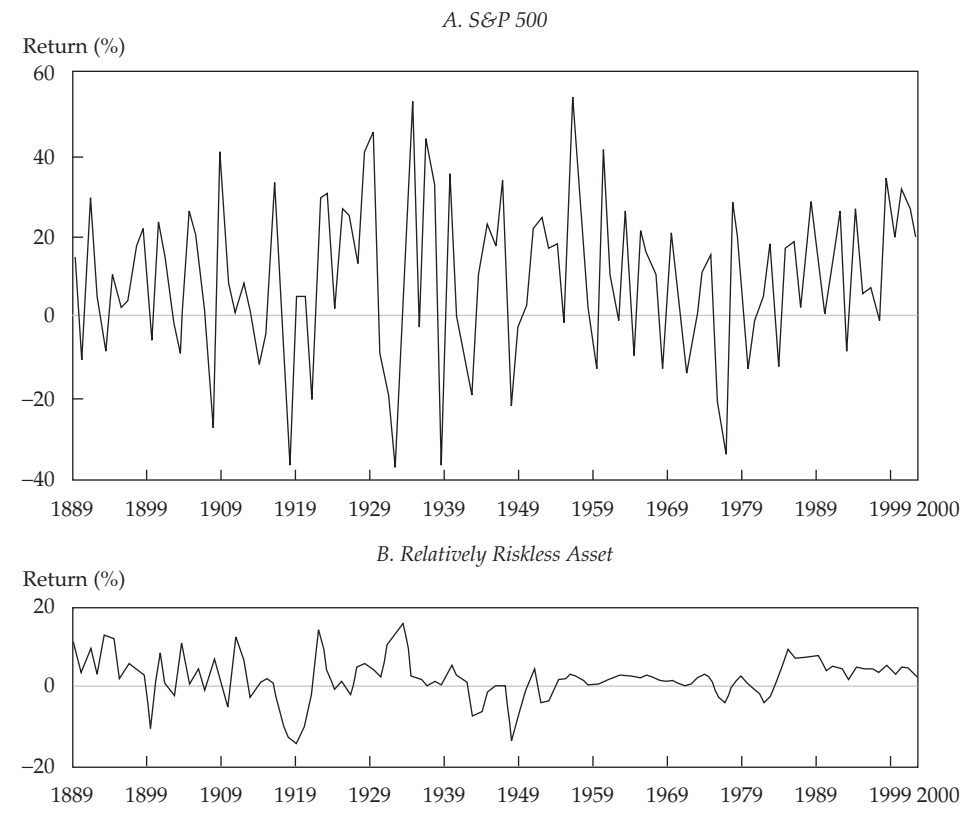
To enhance and deepen our understanding of the risk–return trade-off in the pricing of financial assets, we make a detour into modern asset-pricing theory and look at why different assets yield different rates of return. The *deus ex machina* in asset-pricing theory is that assets are priced in such a way that, *ex ante*, the loss in marginal utility incurred by sacrificing current consumption and buying an asset at a certain price is equal to the expected gain in marginal utility contingent on the anticipated increase in consumption when the asset pays off in the future.

The operative concept here is “incremental loss or gain in well-being because of consumption,” and

the concept should be differentiated from incremental consumption itself. The same amount of consumption may result in different degrees of well-being at different times. (For example, a five-course dinner after a heavy lunch yields considerably less satisfaction than a similar dinner when one is hungry!)

As a consequence, assets that pay off when times are good and consumption levels are high (i.e., when the incremental value of additional consumption is low) are less desirable than those that pay off an equivalent amount when times are bad and additional consumption is both desirable and more highly valued. Thus, assets that pay off when times are good must offer a premium to induce investors to hold them.

Let me illustrate this principle in the context of the standard popular paradigm, the capital asset pricing model. The CAPM postulates a linear relationship between an asset's beta (a measure of

Figure 1. Real Return on S&P 500 and Relatively Riskless Asset, 1889–2000

systematic risk) and expected return. Thus, high-beta stocks yield a high expected rate of return. In the CAPM, good times and bad times are captured by the return on the market. The performance of the market, as captured by a broad-based index, acts as a surrogate indicator for the relevant state of the economy. A high-beta security thus tends to pay off more when the market return is high (when times are good and consumption is plentiful), but as just noted, such a security provides less incremental utility than a security that pays off when consumption is low. It is less valuable to investors and, consequently, sells for less. Thus, assets that pay off in states of low-marginal utility will sell for a lower price than similar assets that pay off in states of high marginal utility. Since rates of return are inversely proportional to asset prices, a high-beta asset will, on average, have a higher rate of return than a low-beta security.

Another perspective on asset pricing emphasizes that economic agents prefer to smooth patterns of consumption over time. Assets that pay off a relatively larger amount at times when consumption is already high “destabilize” these patterns of consumption, whereas assets that pay off when consumption levels are low “smooth out” consumption. Naturally, these latter assets are more

valuable and thus require a lower rate of return to induce investors to hold them. (Insurance policies are a classic example of assets that smooth consumption. Individuals willingly purchase and hold them in spite of their very low rates of return.)

To return to the original question: Are stocks so much riskier than T-bills that a 7 pp differential in their rates of return is justified? What came as a surprise to many economists and researchers in finance was the conclusion of a research paper that Prescott and I wrote in 1979. Stocks and bonds pay off in approximately the same states of nature or economic scenarios, and hence, as argued earlier, they should command approximately the same rate of return. In fact, using standard theory to estimate risk-adjusted returns, we found that stocks, on average, should command, at most, a 1 pp return premium over bills. Because for as long as we had reliable data (about a hundred years), the mean premium on stocks over bills was considerably and consistently higher, we realized that we had a puzzle on our hands. Our paper detailing this research, “The Equity Premium: A Puzzle,” was published in 1985.

To illustrate the puzzle, consider a frictionless economy that has a single representative, “stand-

in" household. This household unit orders its preferences over random consumption paths by

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t U(c_t) \right], 0 < \beta < 1, \quad (1)$$

where

$E_0(\cdot)$ = expectation operator conditional on information available at time zero (which denotes the present time)

β = subjective time discount factor

U = an increasing, continuously differentiable concave utility function

c_t = per capita consumption

The utility function is further restricted to be of the constant relative risk aversion (CRRA) class:

$$U(c, \alpha) = \frac{c^{1-\alpha}}{1-\alpha}, 0 < \alpha < \infty, \quad (2)$$

where the parameter α measures the curvature of the utility function. When $\alpha = 1$, the utility function is defined to be logarithmic, which is the limit of Equation 2 as α approaches 1.

The feature that makes Equation 2 the "preference function of choice" in much of the literature on growth and in Real Business Cycle theory is that it is scale invariant. Although the levels of aggregate variables, such as capital stock, have increased over time, the equilibrium return process is stationary. A second attractive feature is that it is one of only two preference functions that allow for aggregation and a "stand-in" (representative) agent formulation that is independent of the initial distribution of endowments. One disadvantage of this representation is that it links risk preferences with time preferences. With CRRA preferences, agents who like to smooth consumption across various states of nature also prefer to smooth consumption over time; that is, they dislike growth. Specifically, the coefficient of relative risk aversion is the reciprocal of the elasticity of intertemporal substitution. There is no fundamental economic reason why this must be so. I revisit the implications of this issue later, in examining preference structures that do not impose this restriction.³

For this illustration of the puzzle, assume one productive unit that produces in period t output y_t , which is the period dividend. There is one equity share with price p_t (denominated in consumption units) that is competitively traded; it is a claim on the stochastic process $\{y_t\}$.

Consider the intertemporal choice problem of a typical investor at time t . He equates the loss in utility associated with buying one additional unit of equity to the discounted expected utility of the

resulting additional consumption next period. To carry over one additional unit of equity, p_t units of the consumption good must be sacrificed and the resulting loss in utility is $p_t U'(c_t)$. By selling this additional unit of equity next period, $p_{t+1} + y_{t+1}$ additional units of the consumption good can be consumed and $\beta E_t[(p_{t+1} + y_{t+1})U'(c_{t+1})]$ is the expected value of the incremental utility next period. At an optimum, these quantities must be equal. The result is the fundamental pricing relationship:⁴

$$p_t U'(c_t) = \beta E_t[(p_{t+1} + y_{t+1})U'(c_{t+1})]. \quad (3)$$

Equation 3 is used to price both stocks and riskless one-period bonds. For equity,

$$1 = \beta E_t \left[\frac{U'(c_{t+1})}{U'(c_t)} R_{e,t+1} \right], \quad (4)$$

where $R_{e,t+1}$ is equal to $(p_{t+1} + y_{t+1})/p_t$. For the riskless one-period bonds, the relevant pricing expression is

$$1 = \beta E_t \left[\frac{U'(c_{t+1})}{U'(c_t)} R_{f,t+1} \right], \quad (5)$$

The gross rate of return on the riskless asset, R_f , is, by definition,

$$R_{f,t+1} = \frac{1}{q_t}, \quad (6)$$

with q_t as the price of the bond.

We can rewrite Equation 4 as

$$1 = \beta E_t(M_{t+1} R_{e,t+1}), \quad (7)$$

where $M_{t+1} = U'(c_{t+1})/U'(c_t)$. Since $U(c)$ is assumed to be increasing, M_{t+1} is a strictly positive stochastic discount factor. This definition guarantees that the economy will be arbitrage free and the law of one price will hold.

A little algebra demonstrates that the expected gross return on equity is⁵

$$E_t(R_{e,t+1}) = R_{f,t+1} + \text{cov}_t \left\{ \frac{-U'(c_{t+1})}{E_t[U'(c_{t+1})]}, R_{e,t+1} \right\}. \quad (8)$$

The equity premium, $E_t(R_{e,t+1}) - R_{f,t+1}$, can thus be easily computed. Expected asset returns equal the risk-free rate plus a premium for bearing risk, which depends on the covariance of the asset returns with the marginal utility of consumption. Assets that covary positively with consumption—that is, assets that pay off in states when consumption is high and marginal utility is low—command a high premium because these assets "destabilize" consumption.

The question now is: Is the magnitude of the covariance between the assets and the marginal

utility of consumption large enough to justify the observed 6 pp equity premium in U.S. equity markets? In addressing this issue, we make some additional assumptions. Although these assumptions are not necessary and were not part of the original Mehra–Prescott (1985) paper, they facilitate exposition and result in closed-form solutions.⁶ These assumptions are as follows:

- the growth rate of consumption, $x_{t+1} \equiv c_{t+1}/c_t$, is identically and independently distributed (i.i.d.),
- the growth rate of dividends, $z_{t+1} \equiv y_{t+1}/y_t$, is i.i.d., and
- (x_t, z_t) are jointly lognormally distributed.

A consequence of these assumptions is that the gross return on equity, $R_{e,t}$, is i.i.d. and $(x_t, R_{e,t})$ are jointly lognormally distributed.

Substituting $U'(c_t) = c_t^{-\alpha}$ in the fundamental pricing relationship,

$$p_t = \beta E_t \left[(p_{t+1} + y_{t+1}) \frac{U'(c_{t+1})}{U'(c_t)} \right], \quad (9a)$$

we get

$$p_t = \beta E_t [(p_{t+1} + y_{t+1}) x_{t+1}^{-\alpha}]. \quad (9b)$$

It can be easily shown that the expected return on the risky asset is⁷

$$E_t(R_{e,t+1}) = \frac{E_t(z_{t+1})}{\beta E_t(z_{t+1} x_{t+1}^{-\alpha})}. \quad (10)$$

Analogously, the gross return on the riskless asset can be written as

$$R_{f,t+1} = \frac{1}{\beta E_t(x_{t+1}^{-\alpha})}. \quad (11)$$

Because the growth rates of consumption and dividends are assumed to be lognormally distributed,

$$E_t(R_{e,t+1}) = \frac{e^{\mu_z + 1/2\alpha_z^2}}{\beta e^{\mu_x - \alpha\mu_x + 1/2(\alpha_z^2 + \alpha^2\alpha_x^2 - 2\alpha\sigma_{x,z})}} \quad (12a)$$

and

$$\ln E_t(R_{e,t+1}) = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\alpha_x^2 + \alpha\sigma_{x,z}, \quad (12b)$$

where

- $\mu_x = E(\ln x)$
- $\sigma_x^2 = \text{var}(\ln x)$
- $\sigma_{x,z} = \text{cov}(\ln x, \ln z)$
- $\ln x =$ the continuously compounded growth rate of consumption

The other terms involving z and R_e are defined analogously. Furthermore, because the growth rate of consumption is i.i.d., the conditional and unconditional expectations are the same.

Similarly,

$$R_f = \frac{1}{\beta e^{-\alpha\mu_x + 1/2\alpha^2\alpha_x^2}} \quad (13a)$$

and

$$\ln R_f = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\alpha_x^2. \quad (13b)$$

Therefore,

$$\ln E(R_e) - \ln R_f = \alpha\sigma_{x,z}. \quad (14)$$

In this model, it also follows that

$$\ln E(R_e) - \ln R_f = \alpha\sigma_{x,R_e}, \quad (15)$$

where $\sigma_{x,R_e} = \text{cov}(\ln x, \ln R_e)$.

The (log) equity premium in this model is the product of the coefficient of risk aversion and the covariance of the (continuously compounded) growth rate of consumption with the (continuously compounded) return on equity or the growth rate of dividends. If the model equilibrium condition is imposed that $x = z$ (a consequence of which is the restriction that the return on equity be perfectly correlated with the growth rate of consumption), we get

$$\ln E(R_e) - \ln R_f = \alpha\sigma_x^2 \quad (16)$$

and the equity premium is then the product of the coefficient of relative risk aversion, α , and the variance of the growth rate of consumption. As we see later, this variance, σ_x^2 , is 0.00125, so unless α is large, a high equity premium is impossible. The growth rate of consumption just does not vary enough!

Table 4 contains the sample statistics for the U.S. economy for the 1889–1978 period that we reported in Mehra and Prescott (1985).

In our calibration, we are guided by the tenet that model parameters should meet the criteria of

Table 4. U.S. Economy Sample Statistics, 1889–1978

Statistic	Value
Risk-free rate, R_f	1.008
Mean return on equity, $E(R_e)$	1.0698
Mean growth rate of consumption, $E(x)$	1.018
Standard deviation of growth rate of consumption, $\sigma(x)$	0.036
Mean equity premium, $E(R_e) - R_f$	0.0618

cross-model verification. Not only must they be consistent with the observations under consideration, but they should not be *grossly inconsistent* with other observations in growth theory, business cycle theory, labor market behavior, and so on. There is a wealth of evidence from various studies that the coefficient of risk aversion, α , is a small number, certainly less than 10.⁸ We can thus pose the question: If we set risk-aversion coefficient α to be 10 and β to be 0.99, what are the expected rates of return and the risk premium using the parameterization just described?

Using the expressions derived earlier, Appendix A, and Table 4,⁹ we have

$$\begin{aligned} \ln R_f &= -\ln \beta + \alpha \mu_x - \frac{1}{2} \alpha^2 \sigma_x^2 \\ &= 0.120 \end{aligned}$$

or

$$R_f = 1.127,$$

which is a risk-free rate of 12.7 percent!

Because

$$\begin{aligned} \ln E(R_e) &= \ln R_f + \alpha \sigma_x^2 \\ &= 0.132, \end{aligned}$$

we have

$$E(R_e) = 1.141,$$

or a return on equity of 14.1 percent, which implies an equity risk premium of 1.4 pps, far lower than the 6.18 pps historically observed.

Note that in this calculation, I was very liberal in choosing the values for α and β . Most studies indicate a value for α that is close to 2. If I were to pick a lower value for β , the risk-free rate would be even higher and the premium lower. So, the 1.4 pp value represents the maximum equity risk premium that can be obtained, given the constraints on α and β , in this class of models. Because the observed equity premium is more than 6 pps, we have a puzzle on our hands that risk considerations alone cannot account for.

Weil (1989) dubbed the high risk-free rate obtained in the preceding analysis “the risk-free rate puzzle.” The short-term real rate in the United States has averaged less than 1 percent, so the high value of α required to generate the observed equity premium results in an unacceptably high risk-free rate.

The late Fischer Black proposed that $\alpha = 55$ would solve the puzzle.¹⁰ Indeed, it can be shown that the U.S. experience from 1889 through 1978 reported here can be reconciled with $\alpha = 48$ and $\beta = 0.55$. To see this, observe that

$$\begin{aligned} \sigma_x^2 &= \ln \left\{ 1 + \frac{\text{var}(x)}{[E(x)]^2} \right\} \\ &= 0.00125 \end{aligned}$$

and

$$\begin{aligned} \mu_x &= \ln E(x) - \frac{1}{2} \sigma_x^2 \\ &= 0.0172, \end{aligned}$$

which implies that

$$\begin{aligned} \alpha &= \frac{\ln E(R) - \ln R_f}{\sigma_x^2} \\ &= 47.6. \end{aligned}$$

Because

$$\begin{aligned} \ln \beta &= -\ln R_f + \alpha \mu_x - \frac{1}{2} \alpha^2 \sigma_x^2 \\ &= -0.60, \end{aligned}$$

the implication is that β equals 0.55.

Besides postulating an unacceptably high α , another problem is that this is a “knife edge” solution. No other set of parameters will work, and a small change in α will lead to an unacceptably high risk-free rate.¹¹

The relationship $\ln R_f = -\ln \beta + \alpha \mu_x - 1/2 \alpha^2 \sigma_x^2$ shows why an extremely high α can be consistent with a “low” risk-free rate. The term $1/2 \alpha^2 \sigma_x^2$ dominates when α is very large;¹² however, then a small change in the growth rate of consumption will have a large impact on interest rates. This is inconsistent with a cross-country comparison of real risk-free rates and their observed variability. For example, throughout the 1980s, South Korea had a much higher growth rate than the United States but real rates were not appreciably higher. Nor does the risk-free rate vary considerably over time, as would be expected if α were large.

An alternative perspective on the puzzle has been provided by Hansen and Jagannathan (1991). The fundamental pricing equation can be written as

$$E_t(R_{e,t+1}) = R_{f,t+1} - \text{cov}_t \left[\frac{M_{t+1} R_{e,t+1}}{E_t(M_{t+1})} \right]. \quad (17)$$

Equation 17 also holds unconditionally, so

$$E(R_{e,t+1}) = \frac{R_{f,t+1} - \sigma(M_{t+1})\sigma(R_{e,t+1})\rho_{R,M}}{E(M_{t+1})} \quad (18a)$$

or

$$\frac{E(R_{e,t+1}) - R_{f,t+1}}{\sigma(R_{e,t+1})} = -\frac{\sigma(M_{t+1})\rho_{R,M}}{E(M_{t+1})}, \quad (18b)$$

where $\rho_{R,M}$ is the correlation of the return on the security and the stochastic discount factor M . And because $-1 \leq \rho_{R,M} \leq 1$,

$$\left| \frac{E(R_{e,t+1}) - R_{f,t+1}}{\sigma(R_{e,t+1})} \right| \leq \frac{\sigma(M_{t+1})}{E(M_{t+1})}. \quad (19)$$

This inequality is referred to as the “Hansen–Jagannathan lower bound on the pricing kernel.”

For the U.S. economy, the long-term Sharpe ratio, defined as $[E(R_{e,t+1}) - R_{f,t+1}] / \sigma(R_{e,t+1})$, can be calculated to be 0.37. Because $E(M_{t+1})$ is the expected price of a one-period risk-free bond, its value must be close to 1. In fact, for the parameterization discussed earlier, $E(M_{t+1}) = 0.96$ when $\alpha = 2$. Thus, if the Hansen–Jagannathan bound is to be satisfied, the lower bound on the standard deviation for the pricing kernel must be close to 0.3. When this lower bound is calculated in the Mehra–Prescott framework, however, an estimate of 0.002 is obtained for $\sigma(M_{t+1})$, which is off by more than an order of magnitude.

I want to emphasize that the equity premium puzzle is a *quantitative* puzzle; standard theory is consistent with our notion of risk that, on average, stocks should return more than bonds. The puzzle arises from the fact that the quantitative predictions of the theory are an order of magnitude different from what has been historically documented. The puzzle cannot be dismissed lightly because much of our economic intuition is based on the very class of models that fall short so dramatically when confronted with financial data. It underscores the failure of paradigms central to financial and economic modeling to capture the characteristic that appears to make stocks comparatively so risky. Hence, the viability of using this class of models for any quantitative assessment—say, to gauge the welfare implications of alternative stabilization policies—is thrown open to question.

For this reason, over the past 15 years or so, attempts to resolve the puzzle have become a major research impetus in finance and economics. Several generalizations of key features of the Mehra–Prescott (1985) model have been proposed to reconcile observations with theory, including alternative assumptions about preferences (Abel 1990; Benartzi and Thaler 1995; Campbell and Cochrane 1999; Constantinides 1990; Epstein and Zin 1991), modified probability distributions to admit rare but disastrous events (Rietz 1988), survivorship bias (Brown, Goetzmann, and Ross 1995), incomplete markets (Constantinides and Duffie 1996; Heaton and Lucas 1997; Mankiw 1986; Storesletten, Telmer, and Yaron 1999), and market imperfections (Aiyagari and Gertler 1991; Alvarez and Jermann

2000; Bansal and Coleman 1996; Constantinides, Donaldson, and Mehra 2002; Heaton and Lucas 1996; McGrattan and Prescott 2001; Storesletten et al.). None has fully resolved the anomalies. In the next section, I examine some of these efforts to solve the puzzle.¹³

Alternative Preference Structures

The research attempting to solve the equity premium puzzle by modifying preferences can be grouped into two broad approaches—one that calls for modifying the conventional time-and-state-separable utility function and another that incorporates habit formation.

Modifying the Time-and-State-Separable Utility Function.

The analysis in the preceding section shows that the isoelastic (CRRA) preferences used in Mehra and Prescott (1985) can be made consistent with the observed equity premium only if the coefficient of relative risk aversion is implausibly large. A restriction imposed by this class of preferences is that the coefficient of risk aversion is rigidly linked to the elasticity of intertemporal substitution; one is the reciprocal of the other. The implication is that if an individual is averse to variation of consumption in different states at a particular point in time, then she or he will be averse to consumption variation over time. There is no *a priori* reason that this must be so. Since on average, consumption grows over time, the agents in the Mehra–Prescott setup have little incentive to save. The demand for bonds is low and the risk-free rate, as a consequence, is counterfactually high.

To deal with this problem, Epstein and Zin presented a class of preferences, which they termed “generalized expected utility” (GEU), that allows independent parameterization for the coefficient of risk aversion and the elasticity of intertemporal substitution.

In these preferences, utility is recursively defined by

$$U_t = \left[c_t^{1-\rho} + \beta (E_t U_{t+1}^{1-\alpha})^{(1-\rho)/(1-\alpha)} \right]^{1/1-\rho}. \quad (20)$$

The usual isoelastic preferences follow as a special case when $\alpha = \rho$.

The major advantage of this class of models is that a high coefficient of risk aversion, α , does not necessarily imply that agents will want to smooth consumption over time. This modification has the potential to at least resolve the risk-free rate puzzle.

The main difficulty in testing this alternative preference structure stems from the fact that the

counterparts of Equations 3 and 4 when GEU is used depend on the unobserved utility at time $t + 1$, which makes calibration tricky. One would need to make specific assumptions about the consumption process to obtain first-order conditions in terms of observables.

Although Epstein and Zin claimed that their framework offers a solution to the equity premium puzzle, I believe that the proxies they used for the unobservables overstate their claim. The framework does ameliorate the risk-free rate puzzle, but it does not solve the equity premium puzzle.

Habit Formation. A second approach to modifying preferences, initiated by Constantinides (1990), incorporates habit formation. This formulation assumes that utility is affected not only by current consumption but also by past consumption. It captures the notion that utility is a decreasing function of past consumption and marginal utility is an increasing function of past consumption.

Utility is defined as

$$U(c) = E_t \sum_{s=0}^{\infty} \beta^s \frac{(c_{t+s} - \lambda c_{t+s-1})^{1-\alpha}}{1-\alpha}, \lambda > 0, \quad (21)$$

where λ is a parameter that captures the effect of past consumption. This preference ordering makes the agent extremely averse to consumption risk even when the risk aversion is small. For small changes in consumption, changes in marginal utility can be large. So, although this approach cannot solve the equity premium puzzle without invoking extreme aversion to consumption risk, it can address the risk-free rate puzzle because the induced aversion to consumption risk increases the demand for bonds, thereby reducing the risk-free rate.

A modification of the preceding approach is to define utility of consumption relative to average per capita consumption. Abel termed this approach “keeping up with the Joneses.” The idea is that one’s utility depends not on the absolute level of consumption but on how one is doing relative to others. The effect is that, once again, an individual can become extremely sensitive and averse to consumption variation. Equity may have a negative rate of return, which can result in personal consumption falling relative to others’ consumption. Equity thus becomes an undesirable asset relative to bonds. Since average per capita consumption is rising over time, the induced demand for bonds with this modification helps mitigate the risk-free rate puzzle.

An alternate approach, expounded by Campbell and Cochrane, incorporates the possibility of recession—that is, a major economic downturn—as a state variable. In this model, the risk aversion of investors rises dramatically when the chances of a recession increase; thus, the model can generate a high equity premium. Since risk aversion increases precisely when consumption is low, it generates a precautionary demand for bonds that helps lower the risk-free rate. This model is consistent with both consumption and asset market data. Whether investors actually display the huge, time-varying, countercyclical variations in risk aversion postulated in this model is, however, open to question.

In summary, models with habit formation, relative consumption, or subsistence consumption have had success in addressing the risk-free rate puzzle but only limited success with resolving the equity premium puzzle because in these models, effective risk aversion and prudence become improbably large.

Idiosyncratic and Uninsurable Income Risk

In infinite-horizon models, agents, when faced with uninsurable income shocks, dynamically self-insure; agents simply stock up on bonds when times are good and deplete them in bad times, thereby effectively smoothing consumption. Hence, the difference between the equity premium in incomplete markets and in complete markets is small (Heaton and Lucas 1996, 1997). The analysis changes, however, when a shock is permanent. Constantinides and Duffie developed a model that incorporates heterogeneity by capturing the notion that consumers are subject to idiosyncratic income shocks that cannot be insured away.

Simply put, the model recognizes that consumers face the risk of job loss or other major personal disasters that cannot be hedged away or insured against. Equities and related procyclical investments (assets whose payoffs are contingent on the business cycle) exhibit the undesirable feature that they drop in value when the probability of job loss increases—as it does in recessions, for instance. In economic downturns, consumers thus need an extra incentive to hold equities and similar investment instruments; the equity premium can then be rationalized as the added inducement needed to make equities palatable to investors. This model can generate a high risk premium, but whether the required degree of consumption variation can be generated in an economy populated with agents

displaying a relatively low level of risk aversion remains to be seen.

Disaster States and Survivorship Bias

Reitz proposed a solution to the equity premium puzzle that incorporates a small probability of a large drop in consumption. He found that the risk-free rate in such a scenario is much lower than the return on an equity security.

This model requires a 1 in 100 chance of a 25 percent decline in consumption to reconcile the equity premium with a risk-aversion parameter of 10. Such a scenario has not been observed in the United States for the past 100 years (the time for which we have data). Nevertheless, one can evaluate the implications of the model.

One implication is that the real interest rate and the probability of the occurrence of the extreme event move inversely. For example, the perceived probability of a recurrence of a depression was probably high just after WWII but, subsequently, declined over time. If real interest rates had risen significantly as the war years receded, that evidence would support the Reitz hypothesis. Similarly, if the low-probability event precipitating the large decline in consumption is interpreted to be a nuclear war, the perceived probability of such an event has surely varied over the past 100 years. It must have been low before 1945, the first and only year the atom bomb was used, and it must have been higher before the Cuban missile crisis than after it. If real interest rates moved with these sentiments as predicted, that evidence would support Rietz's disaster scenario. But interest rates did not move as predicted.

Another attempt at resolving the puzzle, proposed by Brown et al., focuses on survivorship bias. The central thesis is that the *ex post* measured returns reflect the premium in the United States on a stock market that has successfully weathered the vicissitudes of fluctuating financial fortunes. Many other exchanges have been unsuccessful; hence, the *ex ante* equity premium was low because no one knew *a priori* which exchanges would survive.

For this explanation to work, however, stock and bond markets must be differently affected by financial crises. One reason the effects might not be different is that governments have expropriated much of the real value of nominal debt by the mechanism of unanticipated inflation. Five historical instances come readily to mind: During the post-WWI period of German hyperinflation, holders of bonds denominated in Reich marks lost virtually all of the value invested in those assets.

During the 1920s administration of Henri Poincaré in France, bondholders lost nearly 90 percent of the value invested in nominal debt. In the 1980s, Mexican holders of dollar-denominated debt lost a sizable fraction of the debt's value when the Mexican government, in a period of rapid inflation, converted the debt to pesos and limited the rate at which the funds could be withdrawn. Bondholders in czarist Russia and holders of Chinese debt holdings after the fall of the Nationalist government suffered a similar fate under the new communist regimes.

These examples demonstrate that in times of financial crises, bonds are as likely to lose value as stocks. In every instance when, because of political upheavals or other disasters, trading in equity has been suspended, governments have either reneged on their debt obligations or expropriated much of the real value of nominal debt through unanticipated inflation. Thus, although survivorship bias may have an impact on the levels of the return on both equity and debt, researchers have no evidence that crises affect the returns to stocks and bonds *differentially*. Hence, the equity premium is not affected.

Borrowing Constraints

In models that impose borrowing constraints and transaction costs, these features force investors to hold an inventory of bonds (precautionary demand) to smooth consumption. Hence, in infinite-horizon models with borrowing constraints, agents come close to equalizing their marginal rates of substitution with little effect on the equity premium.¹⁴ Some recent attempts to resolve the equity premium puzzle incorporating both borrowing constraints *and* consumer heterogeneity appear promising. One approach, which departs from the representative agent model, was proposed in Constantinides et al.

The novelty of this paper is the incorporation of a life-cycle feature to study asset pricing. The idea is appealingly simple: The attractiveness of equity as an asset depends on the correlation between consumption and equity income. If equity pays off in states of high marginal utility of consumption, it will command a higher price (and, consequently, a lower rate of return) than if its payoff occurs in states of low marginal utility. Because the marginal utility of consumption varies inversely with consumption, equity will command a high rate of return if it pays off in states when consumption is high, and vice versa.¹⁵

The key insight of the paper is as follows: As the correlation of equity income with consumption

changes over the life cycle of an individual, so does the attractiveness of equity as an asset. Consumption can be decomposed into the sum of wages and equity income. A young person, looking forward, has uncertain future wage and equity income. Furthermore, the correlation of equity income with consumption at this stage is not particularly high, as long as stock and wage income are not highly correlated. This relationship is empirically the case, as documented by Davis and Willen (2000). Equity will thus be a hedge against fluctuations in wages and a “desirable” asset to hold as far as the young are concerned.

The same asset (equity) has a very different characteristic for middle-aged investors. Their wage uncertainty has largely been resolved. Their future retirement wage income is either zero or deterministic, and the innovations (fluctuations) in their consumption occur from fluctuations in equity income. At this stage of the life cycle, equity income is highly correlated with consumption. Consumption is high when equity income is high and equity is no longer a hedge against fluctuations in consumption; hence, for this group, equity requires a higher rate of return.

The characteristics of equity as an asset, therefore, change depending on who the predominant holder of the equity is. Life-cycle considerations thus become crucial for asset pricing. If equity is a “desirable” asset for the marginal investor in the economy, then the observed equity premium will be low relative to an economy in which the marginal investor finds holding equity unattractive. The *deus ex machina* is the stage in the life cycle of the marginal investor.

Constantinides et al. argued that the young, who should (in an economy without frictions and with complete contracting) be holding equity, are effectively shut out of this market because of borrowing constraints. They have low wages; so, ideally, they would like to smooth lifetime consumption by borrowing against future wage income (consuming a part of the loan and investing the rest in higher-returning equity). They are prevented from doing so, however, because human capital alone does not collateralize major loans in modern economies (for reasons of moral hazard and adverse selection).

In the presence of borrowing constraints, equity is thus exclusively priced by middle-aged investors and the equity premium is high. If the borrowing constraint were to be relaxed, the young would borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield would induce the middle-aged to shift their portfolio holdings from equity to bonds. The increase

in the demand for equity by the young and the decrease in the demand for equity by the middle-aged would work in opposite directions. On balance, the effect in the Constantinides et al. model is to increase both the equity and the bond return while simultaneously shrinking the equity premium. Furthermore, the relaxation of the borrowing constraint reduces the net demand for bonds—and the risk-free rate puzzle re-emerges.

Liquidity Premium

Bansal and Coleman developed a monetary model that offers an explanation of the equity premium. In their model, assets other than money play a key role by facilitating transactions, which affects the rate of return they offer in equilibrium.

To motivate the importance of considering the role of a variety of assets in facilitating transactions, Bansal and Coleman argued that, on the margin, the transaction-service return of money relative to interest-bearing checking accounts should be the interest rate paid on these accounts. They estimated this rate, based on the rate offered on NOW accounts for the period they analyzed, to be 6 percent. Because this number is substantial, they suggested that other money-like assets may also implicitly include a transaction-service component in their return. Insofar as T-bills and equity have different service components built into their returns, the Bansal–Coleman argument may offer an explanation for the observed equity premium. In fact, if this service-component differential were about 5 percent, there would be no equity premium puzzle.

This approach can be challenged, however, on three accounts. First, the bulk of T-bill holdings are concentrated in institutions, which do not use them as compensatory balances for checking accounts; thus, it is difficult to accept that they have a significant transaction-service component. Second, the returns on NOW and other interest-bearing accounts have varied over time. For most of the 20th century, checking accounts were not interest bearing, and returns were higher after 1980 than in earlier periods. Yet, contrary to the implications of this model, the equity premium did not diminish in the post-1980 period, when (presumably) the implied transaction-service component was the greatest. Finally, this model implies a significant yield differential between T-bills and long-term government bonds, which (presumably) do not have a significant transaction-service component. However, such a yield differential has not been observed.

Taxes

McGrattan and Prescott proposed an explanation for the equity premium based on changes in tax rates. (An important aspect to keep in mind is that their thesis is not a risk-based explanation. They can account for an equity premium but not as an equity *risk* premium.) McGrattan and Prescott found that, at least in the post-WWII period, the equity premium is not puzzling. They argued that the large reduction in individual income tax rates and the increased opportunity to shelter income from taxation led to a doubling of equity prices between 1960 and 2000. And this increase in equity prices led, in turn, to much higher *ex post* returns on equity than on debt.

This argument can be illustrated by use of a simple one-sector (a corporate sector) model that includes only taxes on corporate distributions and taxes on corporate profits. The authors extended the model to include sufficient details from the U.S. economy—especially in relation to the tax code—to allow them to calibrate the model. They matched up the model with data from the National Income and Product Accounts (NIPA) and the Statistics of Income (SOI) from the Internal Revenue Service.

The model is detailed as follows. Consider a representative-agent economy of infinite life with household preferences defined over consumption and leisure. Each household chooses sequences of consumption and leisure to maximize utility,

$$\max \sum_{t=0}^{\infty} \beta^t U(c_t, l_t), 0 < \beta < 1, \quad (22)$$

subject to the household's budget constraint:

$$\begin{aligned} & \sum_{t=0}^{\infty} p_t [c_t + V_t(s_{t+1} - s_t)] \\ & = \sum_{t=0}^{\infty} p_t [(1 - \tau_{pers}) (d_t s_t + w_t n_t) + \kappa_t], \end{aligned} \quad (23)$$

where

- c_t = per capita consumption at time t
- l_t = fraction of productive time allocated to nonmarket activities, such as leisure
- V_t = price per share
- s_t = number of shares held in period t
- d_t = dividends
- w_t = wages
- κ_t = government transfers
- τ_{pers} = personal taxes

The fraction of time allocated by households to market activities is denoted by $n_t = 1 - l_t$. The budget

constraint represents the condition that the present discounted value of expenditures must be less than or equal to the discounted value of after-tax income. Expenditures of the household are consumption and purchases of stocks, $V_t(s_{t+1} - s_t)$. Income for the households is received from three sources—dividends, wages, and government transfers. Households pay personal taxes, τ_{pers} , on dividend and wage income.

Companies own capital and hire labor to produce output with a constant-return-to-scale production technology,

$$y_t = f(k_{m,t}, k_{u,t}, \xi_t n_t). \quad (24)$$

where

- k_m = tangible assets
- k_u = intangible assets
- n = labor services
- ξ_t = level of technology in period t

Equation 24 thus assumes that companies use both tangible assets and intangible assets to produce output y . Tangible assets include structures, equipment, inventory, and land. Intangible assets are the result of on-the-job training, research and development, and organization and company-specific know-how. In addition to capital, labor services, n , are required. The level of technology in period t , ξ_t , is assumed to grow at rate γ .

Companies choose capital and labor to maximize the present value of dividends. In this framework, therefore, the value of corporate equity, V_t , is

$$V_t = (1 - \tau_{pers}) [k_{m,t+1} + (1 - \tau_{corp}) k_{u,t+1}]. \quad (25)$$

Equation 25 makes clear that a large drop in the personal tax rate with little change in the corporate tax rate, τ_{corp} , will indeed raise the value of equity. However, to show that the market value of U.S. corporate equity relative to GNP, V/y , rises with a decline in the personal tax rate with little change in the corporate tax rate, the capital-to-output ratio *must not* change with changes in τ_{pers} , only with changes in τ_{corp} . McGrattan and Prescott demonstrate this in their paper.

In the United States, corporate income tax rates have changed little since 1960 whereas personal income tax rates have declined significantly. In particular, personal tax rates in the period prior to the tax cuts initiated by the John F. Kennedy administration were considerably higher than in the period after the Tax Reform Act of 1986. McGrattan and Prescott argued that this reduction in personal tax rates was largely unforeseen and that the large unanticipated increase in equity prices had a significant effect on equity returns.

As an illustration, the authors proposed the following hypothetical example. Suppose the tax

rate on dividends were to fall from 50 percent to 0 in 40 years. This fall in the tax rate implies that equity prices would grow by approximately 1.8 pps a year higher than the growth in the GNP. Growth in per capita GNP over the post-WWII period has been 2 percent a year; the growth in population was roughly 1.5 percent in the early part of the postwar period and fell to about 1 percent a year toward the end of the period. Thus, their model would predict an average growth in equity prices of 4.8–5.3 percent a year over the 40-year period. To compute a return, a dividend yield must be added to this price rise. The rise in equity prices implies a fall in the dividend yield because the dividend output ratio stays roughly constant. If the dividend yield were, on average, in the range of 3–4 percent, then equity returns would be in the range of 7.8–9.3 percent, a figure that is of the same magnitude as documented in Table 2 for the post-WWII period.

Additionally, if households are assumed to have a liquidity motive for holding debt, bond returns would be low and the resulting equity premium would be large. The low yield on debt would be reinforced by the additional demand for debt induced by constraints on individuals to hold their retirement assets in debt securities. Indeed, in the first half of the post-WWII period, pension fund assets were almost entirely invested in debt securities because of institutional restrictions on pension fund managers. Before 1974, when ERISA made pension plan fiduciaries personally liable for imprudence or misconduct, fiduciary breaches typically resulted in a loss of tax qualification for the pension fund. Such penalties were likely to be avoidable if pension fund managers held debt assets of various maturities so as to avoid large fund losses and to facilitate the timing of the distribution of benefits.

A potential calibration problem with the McGrattan–Prescott model arises because of the difficulty in identifying the marginal investor. Although the marginal tax rate has dramatically declined in the post-WWII period, the same cannot be said about the rates that apply to the marginal investor unless the marginal investor can be identified. The marginal tax rate and the rate that applies to the marginal investor may be quite different, and of course, it is the rate that applies to the marginal investor that is relevant for pricing. To estimate the tax rate that applies to the marginal investor, McGrattan and Prescott computed a weighted-average rate, averaged across income groups, for each year in the 1947–96 period. The accuracy of this estimate, however, is open to question.

The McGrattan–Prescott model predicts that, eventually, as the dividend yield falls and tax rates

level out, everything else being equal, equity returns will decline for the following reason. The real before-tax return on equity is the sum of three returns—the dividend yield, the anticipated capital gain, and the unanticipated capital gain. The dividend yield has been high (more than 3 percent) for much of the post-WWII period because high tax rates have implied a low price of equity. Recently, it has declined, and it is presently just over 1 percent. The anticipated capital gain is the growth rate of productive assets, which is roughly 3 percent. This growth rate has not changed. The unanticipated capital gain is the growth in the price of equity as a result of unanticipated changes in tax rates. This growth rate *has* changed—falling from a range of 1.5–2.0 percent to 0. Adding these rates leads to about an expected (3 + 3 + 2 percent) 8 percent real, before-tax stock return in the early post-WWII period and, barring any further unexpected changes in tax rates, a (1 + 3 + 0 percent) 4 percent return in the future.

No Premium?

An alternative point of view held by a group of academicians and professionals is that currently there is no equity premium and, by implication, no equity premium puzzle. To address these claims, two different interpretations of the term “equity premium” must be distinguished. One is the *ex post* or realized equity premium. This figure is the actual, historically observed difference between the return on the market, as captured by a stock index, and the risk-free rate, as proxied by the return on government bills. This premium is what Prescott and I addressed in our 1985 paper. The other (related) concept is the *ex ante* equity premium. This figure is a forward-looking measure of the premium—that is, the equity premium that is *expected* to prevail in the future or the conditional equity premium given the current state of the economy. For example, after a bull market, stock valuations are high relative to fundamentals, the market has risen sharply, and the *ex post* (realized) equity premium is high, but such a time is precisely when the *ex ante* (expected) equity premium is likely to be low. Conversely, after a major downward correction, the realized premium will be low whereas the expected premium is likely to be high. This relationship should not come as a surprise, because returns to stock have been documented to be mean reverting.

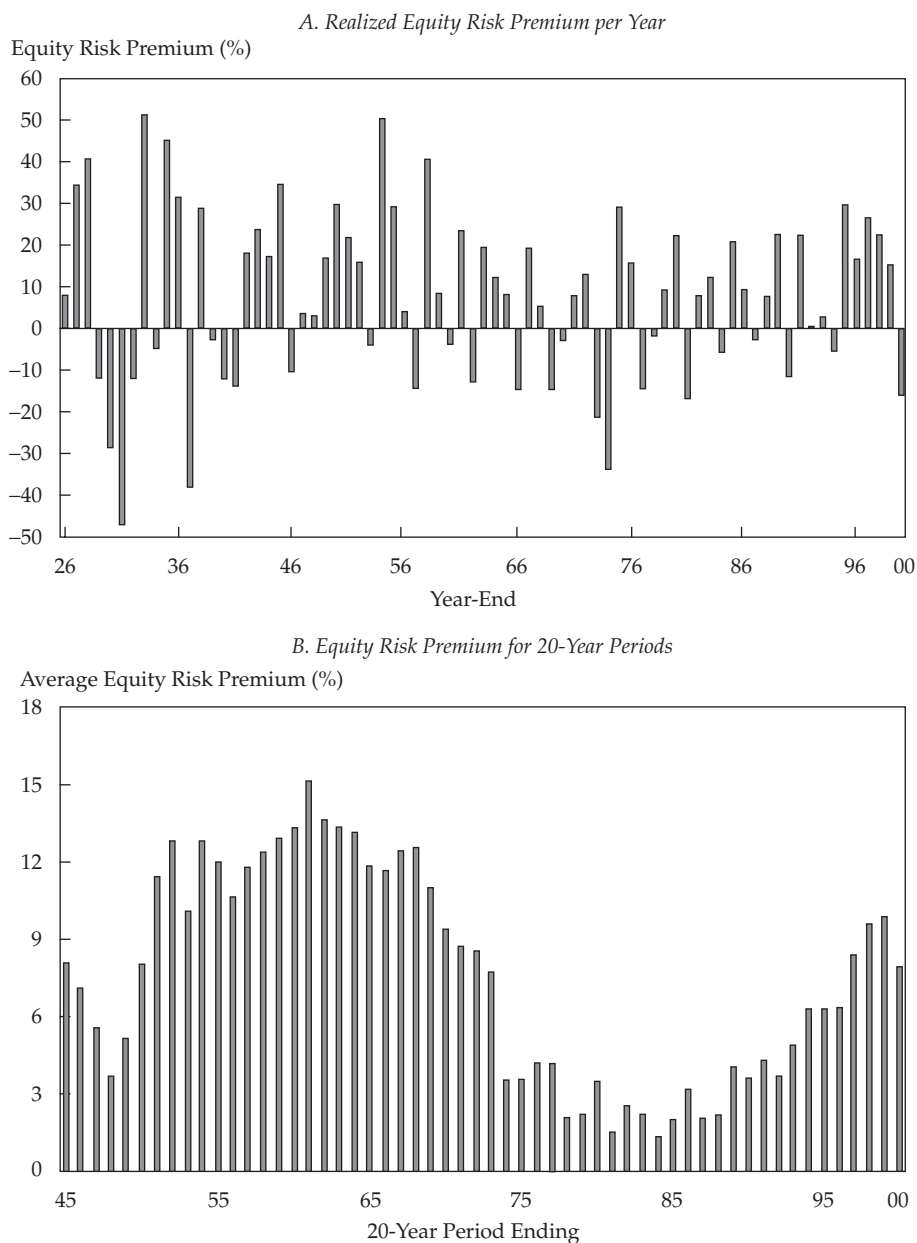
Which of these interpretations of the equity premium is relevant for an investment advisor? Clearly, the choice depends on the planning horizon. The equity premium documented in our 1985

paper reflects very long investment horizons. It has little to do with what the premium is going to be in the next couple of years. The *ex post* equity premium is the realization of a stochastic process over a certain period, and it has varied considerably over time.

Furthermore, the variation in the realized premium depends on the time horizon over which it is measured. In some periods, it has even been negative, as illustrated in Figure 2.

Market watchers and other professionals who are interested in short-term investment planning will wish to project the conditional, expected equity premium over their planning horizon. This task is by no means simple. Even if the equity premium in current market conditions is small (and the general consensus is that it is), it does not imply that either the historical premium was too high or that the equity premium has diminished.

Figure 2. Risk Premium, 1926–2000



Source: Ibbotson (2001).

The data used to document the equity premium over the past 100 years are as good an economic data set as analysts have, and a span of 100 years is a long series when it comes to economic data. Before the equity premium is dismissed, not only do researchers need to understand the observed phenomena, but they also need a plausible explanation as to why the future is likely to be any different from the past. In the absence of this explanation, and on the basis of what is currently known, I make the following claim: Over the long term, the equity premium is likely to be similar to what it has been in the past and returns to investment in equity will continue to substantially dominate returns to investment in T-bills for investors with a long planning horizon.

I thank George Constantinides, Sanjiv Das, John Donaldson, Mark Rubinstein, and especially, Edward Prescott for helpful discussions and Chaitanya Mehra for editorial assistance. This research was supported by a grant from the Academic Senate of the University of California.

Appendix A. Lognormal Properties and Derivations

Some facts about the lognormal distribution follow:

1. If $\ln z \sim N(\mu_z, \sigma_z^2)$, then $a \ln z \sim N(a\mu_z, a^2\sigma_z^2)$.
2. $E(z^a) = E[\exp(a \ln z)] = \exp\left(a\mu_z + \frac{1}{2}a^2\sigma_z^2\right)$.
3. $a \ln z + b \ln x \sim N(a\mu_z + b\mu_x, a^2\sigma_z^2 + b^2\sigma_x^2 + 2ab\rho\sigma_x\sigma_z)$, where $\rho = \text{cor}(\ln x, \ln z)$.
4. $E(z^a x^b) = \exp\left[a\mu_z + b\mu_x + \frac{1}{2}(a^2\sigma_z^2 + b^2\sigma_x^2)\right]$.
5. If $x = z$, then $E(z^a x^b) = E(x^{a+b}) = \exp\left[(a+b)\mu_x + \frac{1}{2}(a+b)^2\sigma_x^2\right]$.
6. $\text{var}(x) = E(x^2) - [E(x)]^2 = \exp(2\mu_x + 2\sigma_x^2) - \exp(2\mu_x + \sigma_x^2) = \exp(2\mu_x + \sigma_x^2)[\exp(\sigma_x^2) - 1] = [E(x)]^2[\exp(\sigma_x^2) - 1]$.
7. $\exp(\sigma_x^2) = 1 + \frac{\text{var}(x)}{[E(x)]^2}$; hence,

$$\sigma_x^2 = \ln\left\{1 + \frac{\text{var}(x)}{[E(x)]^2}\right\}.$$

$$8. \quad \ln E(x) = \mu_x + \frac{1}{2}\sigma_x^2;$$

hence,

$$\mu_x = \ln E(x) - \frac{1}{2}\sigma_x^2.$$

Derivation of Equation 8. Expanding Equation 7 results in

$$\beta E_t\left[\frac{U'(c_{t+1})}{U'(c_t)}\right]E_t(R_{e,t+1}) + \beta \text{cov}_t\left[\frac{U'(c_{t+1}), R_{e,t+1}}{U'(c_t)}\right] = 1.$$

Substituting for $R_{f,t+1}$ from the following equation,

$$1 = \beta E_t\left[\frac{U'(c_{t+1})}{U'(c_t)}\right]R_{f,t+1},$$

results in

$$E_t(R_{e,t+1}) = R_{f,t+1} + \text{cov}_t\left\{\frac{-U'(c_{t+1}), R_{e,t+1}}{E_t[U'(c_{t+1})]}\right\},$$

which is Equation 8 in the article.

Derivation of Equations 10–13. Because p_t is homogeneous of degree 1 in y , we can represent it as

$$p_t = wy_t.$$

Substituting for $U'(c_t) = c_t^{-\alpha}$ and p_t in the following fundamental pricing relationship,

$$p_t = \beta E_t\left[(p_{t+1} + y_{t+1})\frac{U'(c_{t+1})}{U'(c_t)}\right],$$

results in

$$wy_t = \beta E_t[(wy_{t+1} + y_{t+1})x_{t+1}^{-\alpha}];$$

hence,

$$w = \beta E_t[(w+1)z_{t+1}x_{t+1}^{-\alpha}]$$

or

$$w = \frac{\beta E_t(z_{t+1}x_{t+1}^{-\alpha})}{1 - \beta E_t(z_{t+1}x_{t+1}^{-\alpha})}.$$

By definition, $R_{e,t+1}$, the gross rate of return on equity, is

$$R_{e,t+1} = \frac{p_{t+1} + y_{t+1}}{p_t}.$$

Substituting for p_t , we get

$$R_{e,t+1} = \left(\frac{w+1}{w}\right)\left(\frac{y_{t+1}}{y_t}\right)$$

$$= \frac{w+1}{w}z_{t+1}$$

or

$$E_t(R_{e,t+1}) = \frac{w+1}{w}E_t(z_{t+1}).$$

Because

$$\frac{w+1}{w} = \frac{1}{\beta E_t(z_{t+1}x_{t+1}^{-\alpha})},$$

we have

$$E_t(R_{e,t+1}) = \frac{E_t(z_{t+1})}{\beta E_t(z_{t+1}x_{t+1}^{-\alpha})},$$

which is Equation 10 in the text.

Analogously, the gross return on the riskless asset can be written as

$$R_{f,t+1} = \left(\frac{1}{\beta}\right)\left[\frac{1}{E_t(x_{t+1}^{-\alpha})}\right].$$

Using the lognormal properties of Fact 2 and Fact 4, we get

$$E_t(R_{e,t+1}) = \frac{e^{\mu_z+1/2\sigma_z^2}}{\beta e^{\mu_x-\alpha\mu_x+1/2(\sigma_z^2+\alpha^2\sigma_x^2-2\alpha\sigma_{x,z})}}$$

and

$$R_f = \frac{1}{\beta e^{-\alpha\mu_x+1/2\alpha^2\sigma_x^2}}.$$

Taking logs on both sides results in

$$\ln E_t(R_{e,t+1}) = -\ln \beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 + \alpha\sigma_{x,z}$$

and

$$\ln R_f = -\ln \beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2.$$

Notes

1. For an elaboration of the issues presented here, see Mehra (2002) and Mehra and Prescott (forthcoming 2003); some sections of this article closely follow the exposition in that paper. For current approaches to solving the equity risk premium puzzle, see the presentations and discussions at www.aimrpubs.org/ap/home.html from AIMR's *Equity Risk Premium Forum*.
2. The calculations in Table 3 assume that all payments to the underlying asset, such as dividend payments to stock and interest payments to bonds, were reinvested and that no taxes were paid.
3. See Epstein and Zin (1991) and Weil (1989).
4. Versions of this expression can be found in Rubinstein (1976), Lucas (1978), Breeden (1979), Prescott and Mehra (1980), and Donaldson and Mehra (1984), among others. Excellent textbook treatments of asset pricing are available in Cochrane (2001), Danthine and Donaldson (2001), Duffie (2001), and LeRoy and Werner (2001).
5. The derivation is given in Appendix A.
6. The exposition in the text is based on Abel (1988) and his unpublished notes. I thank him for sharing them with me.
7. The derivation of Equations 10–13 is given in Appendix A.
8. A number of these studies are documented in Mehra and Prescott (1985).
9. For instance, to get σ_x^2 , we use Fact 7 and plug in $\text{var}(x)$ and $E(x)$ from Table 4. The properties of the lognormal distribution are documented in Appendix A.
10. Private communication, 1981.
11. An alternate approach is to experiment with negative time preferences; however, there seems to be no empirical evidence that agents do have such preferences.
12. Kandel and Stambaugh (1991) suggested this approach.
13. See also the excellent surveys by Kocherlakota (1996), Cochrane (1997), and Campbell (forthcoming 2003).
14. This is true unless the supply of bonds is unrealistically low. See Aiyagari and Gertler.
15. This argument is precisely the reason high-beta stocks in the simple CAPM framework have a high rate of return. In that model, the return on the market is a proxy for consumption. High-beta stocks pay off when the market return is high (i.e., when marginal utility is low); hence, their price is (relatively) low and their rate of return, relatively high.

References

- Abel, A.B. 1988. "Stock Prices under Time Varying Dividend Risk: An Exact Solution in an Infinite-Horizon General Equilibrium Model." *Journal of Monetary Economics*, vol. 22, no. 3 (November):375–394.
- . 1990. "Asset Prices under Habit Formation and Catching Up with the Joneses." *American Economic Review*, vol. 80, no. 2 (May):38–42.
- Aiyagari, S.R., and M. Gertler. 1991. "Asset Returns with Transactions Costs and Uninsured Individual Risk." *Journal of Monetary Economics*, vol. 27, no. 3 (June):311–331.
- Alvarez, F., and U. Jermann. 2000. Efficiency, Equilibrium, and Asset Pricing with Risk of Default. *Econometrica*, vol. 68, no. 4 (July):775–797.

- Bansal, R., and J.W. Coleman. 1996. "A Monetary Explanation of the Equity Premium, Term Premium, and Risk-Free Rate Puzzles." *Journal of Political Economy*, vol. 104, no. 6 (December):1135–71.
- Barberis, N., M. Huang, and T. Santos. 2001. "Prospect Theory and Asset Prices." *Quarterly Journal of Economics*, vol. 116, no. 1 (February):1–53.
- Benartzi, S., and R.H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *Quarterly Journal of Economics*, vol. 110, no. 1 (February):73–92.
- Breeden, D. 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., W. Goetzmann, and S. Ross. 1995. "Survival." *Journal of Finance*, vol. 50, no. 2 (June):853–873.
- Campbell, J.Y. Forthcoming 2003. "Consumption Based Asset Pricing." In *Handbook of the Economics of Finance*. Edited by G.M. Constantinides, M. Harris, and R. Stulz. Amsterdam, Netherlands: North-Holland.
- Campbell, J.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.
- Cochrane, J.H. 1997. "Where Is the Market Going? Uncertain Facts and Novel Theories." *Economic Perspectives*, vol. 21, no. 6 (November):3–37.
- . 2001. *Asset Pricing*. Princeton, NJ: Princeton University Press.
- Constantinides, G.M. 1990. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy*, vol. 98, no. 3 (June):519–543.
- . 2002. "Rational Asset Prices." *Journal of Finance*, vol. 57, no. 4 (August):1567–91.
- Constantinides, G.M., and D. Duffie. 1996. "Asset Pricing with Heterogeneous Consumers." *Journal of Political Economy*, vol. 104, no. 2 (April):219–240.
- Constantinides, G.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.
- Danthine, J.-P., and J.B. Donaldson. 2001. *Intermediate Financial Theory*. Upper Saddle River, NJ: Prentice Hall.
- Davis, S.J., and P. Willen. 2000. "Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits." Working paper. Chicago, IL: University of Chicago.
- Donaldson, J.B., and R. Mehra. 1984. "Comparative Dynamics of an Equilibrium Intertemporal Asset Pricing Model." *Review of Economic Studies*, vol. 51, no. 1 (July):491–508.
- Duffie, D. 2001. *Dynamic Asset Pricing Theory*. Princeton, NJ: Princeton University Press.
- Epstein, L.G., and S.E. Zin. 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.
- Hansen, L.P., and R. Jagannathan. 1991. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy*, vol. 99, no. 2 (April):225–262.
- Heaton, J., and D.J. Lucas. 1996. "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing." *Journal of Political Economy*, vol. 104, no. 3 (June):443–487.
- . 1997. "Market Frictions, Savings Behavior and Portfolio Choice." *Journal of Macroeconomic Dynamics*, vol. 1, no. 1:76–101.
- Ibbotson Associates. 2001. "Stocks, Bonds, Bills, and Inflation." 2000 Yearbook. Chicago, IL: Ibbotson Associates.
- Kandel, S., and R.F. Stambaugh. 1991. "Asset Returns and Intertemporal Preferences." *Journal of Monetary Economics*, vol. 27, no. 1 (February): 39–71.
- Kocherlakota, N.R. 1996. "The Equity Premium: It's Still a Puzzle." *Journal of Economic Literature*, vol. 34, no. 1 (March):42–71.
- LeRoy, S.-F., and J. Werner. 2001. *Principles of Financial Economics*. Cambridge, U.K.: Cambridge University Press.
- Lucas, R.E., Jr. 1978. "Asset Prices in an Exchange Economy." *Econometrica*, vol. 46, no. 6 (November):1429–45.
- Mankiw, N.G. 1986. "The Equity Premium and the Concentration of Aggregate Shocks." *Journal of Financial Economics*, vol. 17, no. 1 (September):211–219.
- McGrattan, E.R., and E.C. Prescott. 2001. "Taxes, Regulations, and Asset Prices." Working Paper No. 610, Federal Reserve Bank of Minneapolis.
- Mehra, R. 2002. "Current Estimates and Prospects for Change II." In *Equity Risk Premium Forum*. Charlottesville, VA: AIMR (www.aimrpubs.org/ap/issues/v2002n1/toc.html).
- Mehra, R., and E.C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics*, vol. 15, no. 2 (March):145–161.
- . 1988. "The Equity Premium: A Solution?" *Journal of Monetary Economics*, vol. 22, no. 1 (July):133–136.
- . Forthcoming 2003. "The Equity Premium in Retrospect." In *Handbook of the Economics of Finance*. Edited by G.M. Constantinides, M. Harris, and R. Stulz. Amsterdam, Netherlands: North-Holland.
- Prescott, E.C., and R. Mehra. 1980. "Recursive Competitive Equilibrium: The Case of Homogeneous Households." *Econometrica*, vol. 48, no. 6 (September):1365–79.
- Rietz, T.A. 1988. "The Equity Risk Premium: A Solution." *Journal of Monetary Economics*, vol. 22, no. 1 (July):117–131.
- Rubinstein, M. 1976. "The Valuation of Uncertain Income Streams and the Pricing of Options." *Bell Journal of Economics*, vol. 7, no. 2 (Autumn):407–425.
- Siegel, J. 1998. *Stocks for the Long Run*. 2nd ed. New York: Irwin.
- Storesletten, K., C.I. Telmer, and A. Yaron. 1999. "Asset Pricing with Idiosyncratic Risk and Overlapping Generations." Working paper. Pittsburgh: Carnegie Mellon University.
- Weil, P. 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle." *Journal of Monetary Economics*, vol. 24, no. 3 (November):401–421.