THE EQUITY PREMIUM: WHY IS IT A PUZZLE?

Rajnish Mehra

University of California, Santa Barbara
and
University of Chicago

Current draft: May 6, 2001

I thank George Constantinides, John Donaldson and Edward Prescott for helpful discussions. This research was support by a grant from the Academic Senate of the University of California. E-mail: mehra@econ.ucsb.edu.
Historical data provides us with a wealth of evidence documenting that for over a century stock returns have been considerably higher than those for T-bills. The average annual real return, (that is to say, the inflation adjusted return) on the U.S. stock market over the last hundred and ten years has been about 7.9 percent. Over the same period, the return on a relatively riskless security was a paltry 1.0 percent. The difference between these two returns, 6.9 percent, is termed the “equity premium”. This statistical difference is even more pronounced over the post war period, with the premium of stock returns over bonds being almost 8 percent.

Furthermore, this pattern of excess returns to equity holdings is not unique to U.S. capital markets. Equity returns compared to the return to debt holdings in other countries also exhibit this historical regularity. The annual return on the British stock market was 5.7 percent over the post war period, an impressive 4.6 percent premium over the average bond return of 1.1 percent. Similar statistical differentials are documented for France, Germany, Italy and Spain.

Jeremy Siegel (1998) has analyzed data on U.S. stock and bond returns going back to 1802 and found a similar though somewhat smaller premium in place for the past two hundred years.

The table below summarizes the foregoing data.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Mean % real return on a market index</th>
<th>Mean % real return on a relatively riskless security</th>
<th>Mean % risk premium</th>
</tr>
</thead>
<tbody>
<tr>
<td>1802-1998</td>
<td>7.0</td>
<td>2.9</td>
<td>4.1</td>
</tr>
<tr>
<td>1889-1978</td>
<td>6.98</td>
<td>0.80</td>
<td>6.18</td>
</tr>
<tr>
<td>1889-2000</td>
<td>7.9</td>
<td>1.0</td>
<td>6.9</td>
</tr>
<tr>
<td>1926-2000</td>
<td>8.7</td>
<td>0.7</td>
<td>8.0</td>
</tr>
<tr>
<td>1947-2000</td>
<td>8.4</td>
<td>0.6</td>
<td>7.8</td>
</tr>
</tbody>
</table>

Source: 1802-1998 from Siegel (1998), 1889-1978 from Mehra & Prescott (1985) and the rest are the author’s estimates.

The dramatic investment implications of these differential rate of return can be seen from the following table, which maps the capital appreciation of $1 invested in
different assets from 1802 - 1997 and from 1926-2000.

<table>
<thead>
<tr>
<th>Investment period</th>
<th>Stocks</th>
<th>T-bills</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Real</td>
<td>Nominal</td>
</tr>
<tr>
<td>1802-1997</td>
<td>$558,945</td>
<td>$7,470,000</td>
</tr>
<tr>
<td>1926-2000</td>
<td>$266.47</td>
<td>$2,586.52</td>
</tr>
</tbody>
</table>

Source: Ibbotson (2001) and Siegel (1998)

As the table illustrates, $1 invested in a diversified stock index yields an ending wealth of $558,945 versus a value of $276, in real terms, for $1 invested in a portfolio of T-bills for the period 1802-1997. While for the 75-year period 1926-2000 the corresponding values are $266.47 and $1.71. We assume that all payments to the underlying asset, such as dividend payments to stock and interest payments to bonds are reinvested and that there are no taxes paid.

This long-term perspective underscores the remarkable wealth building potential of the equity premium. It should come as no surprise therefore, that the equity premium is of central importance in portfolio allocation decisions, estimates of the cost of capital and is front and center in the current debate about the advantages of investing Social Security funds in the stock market.

Why have stocks been such an attractive investment relative to bonds? Why has the rate of return on stocks been higher than on relatively risk free assets? One intuitive answer is that since stocks are ‘riskier’ than bonds, investors require a larger premium for bearing this additional risk; and indeed, the standard deviation of the returns to stocks (about 20% per annum historically) is larger than that of the returns to T-bills (about 4% per annum), so, obviously they are considerably more risky than bills! But are they?
The figures below show the annual real rate of return on the S&P 500 index and a relatively risk free security over the period 1889-2000.

To enhance and deepen our understanding of the risk-return trade-off in the pricing of financial assets, we make a brief detour into modern asset pricing theory and look at why different assets yield different rates of return. The deux ex machina of this theory is that assets are priced such 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. (This is represented in equations 1 and 2 below).

The operative emphasis here is the incremental loss or gain of well-being due to consumption and should be differentiated from incremental consumption. This is because the same amount of consumption may result in different degrees of well-being at different times. (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.

Let us illustrate this principle in the context of the standard, popular paradigm, the Capital Asset Pricing Model (CAPM) as an example. The model postulates a linear relationship between an asset’s ‘beta’, a measure of systematic risk, and expected return. Thus, high beta stocks yield a high expected rate of return. That is so because 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 tends to pay off more when the market return is high, that is, when times are good and consumption is plentiful; as discussed earlier, such a security provides less incremental well-being than a security that pays off when consumption is low, is less valuable and consequently sells for less. To use the terminology of modern asset pricing theory, assets that pay off in states of low marginal utility will sell for a lower price than similar asset that pay off in states of high marginal utility. Since rates of return are inversely proportional to asset prices, the latter class of assets will, on average, give a lower rate of return than the former.
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, the latter are more valuable and thus require a lower rate of return to induce investors to hold these assets. (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 that much more riskier than bills so as to justify a 7% differential in their rates of return?

What came as a surprise to many economists and researchers in finance was the conclusion of a research paper that I co-authored with Edward Prescott 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, we found, using standard theory to estimate risk-adjusted returns, that stocks on average should command, at most, a 1% return premium over bills. Since, 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. It took us six more years to convince a skeptical profession and for our paper “The Equity Premium: A Puzzle” to be published. (Mehra and Prescott (1985).)

To illustrate the puzzle we start with the fundamental asset pricing relation (Lucas (1978) equation 6) and use it to price stocks and riskless one period bonds.

For stocks we have

\[ p_t u'(c_t) = \beta E_t \{ (p_{t+1} + c_{t+1}) u'(c_{t+1}) \} \]  
(1)

and for the riskless one period bonds the relevant expression is

\[ q_t u(c_t) = \beta E_t \{ u'(c_{t+1}) \} \]  
(2)

Here \( p \) is the price of equity, \( q \) the price of the riskless bond, \( c \) represents consumption, \( \beta \) is the discount factor and \( u(\cdot) \) the period utility function.
The gross rate of return on the riskless asset is by definition:

$$R_{f,t+1} = \frac{1}{q_t}$$

We make the following assumptions:

a) \( u(c) = \frac{c^{1-\alpha} - 1}{1 - \alpha} \); that is, the utility function is of the Constant Relative Risk Aversion type with \( \alpha \) as the coefficient of relative risk aversion.

b) the growth rate of consumption \( x_{t+1} = \frac{c_{t+1}}{c_t} \) is i.i.d.

c) the growth rate of consumption is log normally distributed.

From the fundamental pricing relation:

$$p_t u'(c_t) = \beta E_t \{(p_{t+1} + c_{t+1}) u'(c_{t+1})\}$$

or

$$p_t = \beta E_t \{(p_{t+1} + c_{t+1}) \frac{u'(c_{t+1})}{u'(c_t)} \}$$

Since \( u'(c_t) = c_t^{-\alpha} \)

$$p_t = \beta E_t \{(p_{t+1} + c_{t+1}) x_t^{-\alpha} \}$$

Then, the solution will be of the form:

$$p_t = w c_t$$

$$w c_t = \beta E_t \{(w c_{t+1} + c_{t+1}) x_t^{-\alpha} \}$$

$$w = \beta E_t \{(w + 1) x_t^{1-\alpha} \}$$

or

$$w = \frac{\beta E_t\{x_t^{1-\alpha}\}}{1 - \beta E_t\{x_t^{-\alpha}\}}$$

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

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

Substituting for \( p_t \) we get

$$R_{e,t+1} = \frac{(w + 1)}{w} \cdot \frac{c_{t+1}}{c_t} = \frac{w + 1}{w} \cdot x_{t+1}$$
or
\[ E_t \{ R_{e, t+1} \} = \frac{w + 1}{w} \cdot E_t \{ x_{t+1} \} \]

since
\[ \frac{w + 1}{w} = \frac{1}{\beta E_t \{ x_{t+1}^{1-\alpha} \}} \]

we have
\[ E_t \{ R_{e, t+1} \} = \frac{E_t \{ x_{t+1} \}}{\beta E_t \{ x_{t+1}^{1-\alpha} \}} \] (3)

Analogously, the gross return on the riskless asset can be written as
\[ R_{f, t+1} = \frac{1}{\beta} \frac{1}{E_t \{ x_{t+1}^{1-\alpha} \}} \] (4)

Next we get an expression for \( \frac{E_t \{ R_e \}}{R_f} \)
\[ \frac{E_t \{ R_e \}}{R_f} = \frac{E \{ x \} E \{ x^{1-\alpha} \}}{E \{ x^{1-\alpha} \}} \] (5)

Since we have assumed the growth rate of consumption to be log normally distributed, i.e.,
\[ \ln x \sim N (\mu_x, \sigma_x^2) \]

we get
\[ \frac{E_t \{ R_e \}}{R_f} = \exp\{ \alpha \sigma_x^2 \} \] (6)

\[ \therefore \ln E \{ R_e \} - \ln R_f = \alpha \sigma_x^2 \]

From (4) we get
\[ \ln R_f = -\ln \beta - \ln E \{ x^{1-\alpha} \} \]

\[ \therefore \ln \beta = -\ln R_f - \ln \{ \exp (-\alpha \mu_x + 1/2 \alpha^2 \sigma_x^2) \} \]

or
\[ \ln \beta = -\ln R_f - \alpha \mu_x - 1/2 \alpha^2 \sigma_x^2 \]

Mehra & Prescott (1985) report the following sample statistics for the U.S. economy over the period 1889-1978:

risk free rate = 0.008
mean return on equity = 0.0698
mean growth rate of consumption = 0.0183
standard deviation of the growth rate of consumption = 0.0357
mean equity premium = 0.0618
If we set the risk aversion coefficient $\alpha$ to be 10 and $\beta$ to be 0.99 what are the expected rates of return and the risk premium using the parameterization above?

Using the expressions derived earlier we have

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

or

$$R_f = 1.1318$$

that is, a risk free rate of 13.18%!

Since

$$\ln E\{R_e\} = \ln R_f + \alpha \sigma_x^2$$

we have

$$E\{R_e\} = 1.1458$$

or a return on equity of 14.58%. This implies a risk premium of 1.40%, far lower than the 6.18% historically observed. In this calculation we have been very liberal in choosing the values for $\alpha$ and $\beta$. Most studies indicate a value for $\alpha$ that is close to 3. If we pick a lower value for $\beta$ the riskfree rate will be even higher and the premium lower. So the 1.4% value represents the maximum equity premium that can be obtained in this class of models. Since the observed premium is over 6% we have a puzzle on our hands.

I would like to emphasize that the equity premium puzzle is a quantitative puzzle in that 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, since 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, for instance, to gauge the welfare implications of alternative stabilization policies, is thrown open to question.

For this reason, over the last 15 years or so, attempts to resolve the puzzle have
become a major research impetus in finance and economics. I shall not review this vast literature here but refer the reader to the surveys by Narayana Kocherlakota (1996) and by John Cochrane (1997).

Briefly, the majority of the research efforts to resolve the puzzle have either postulated modifications of standard utility functions to model investors as highly risk averse or have proposed explanations based on market imperfections, transactions costs, potential disaster states, selection bias, and the inability to insure against risk and disaster scenarios.

Some recent attempts to resolve the puzzle appear promising. The first incorporates the possibility of recession, that is to say, a major economic downturn as a state variable. It may be the case that the risk aversion of investors rises dramatically when the chances of a recession become larger. This is the approach expounded by Campbell and Cochrane (1999). The second approach incorporates consumer heterogeneity and departs from the representative agent model. This approach has been proposed Constantinides, Donaldson and Mehra (2001).

Constantinides et al model an economy as consisting of three overlapping generations, the young, the middle-aged and the old, where each cohort, by their consumption and investment decisions affects the demand for and thus the prices of assets in the economy. The young are restricted from participating in the stock market because they face a binding borrowing constraint. Hence stocks are priced not by the young, for whom it is an attractive asset, but by the middle aged, for whom stocks tend to be far less attractive. This is because for the young agents stocks and wages are not highly correlated and hence stocks are a potential hedge against wage fluctuations. The middle aged, however, have no significant future wages, and thus the fluctuations in their consumption arises directly from fluctuations in the value of their holdings of stocks. The latter two quantities are thus highly correlated.

Another form of heterogeneity, proposed by Constantinides and Duffie (1996), captures the notion that consumers are subject to idiosyncratic income shocks that cannot be insured away. Simply put, consumers face the risk of job loss, or other major personal disasters that cannot be hedged away or insured against. Thus, equities and related procyclical investments exhibit the undesirable feature that they drop in value when the
probability of job-loss increases, as it does, for instance, in recessions. In economic downturns, consumers need an extra incentive to hold equities and other similar investment instruments. The equity premium is thus rationalized as the added inducement needed to make equities palatable to investors.

There is also an alternative point of view, held by a group of academicians and professionals who claim that at present there is no equity premium and by implication no equity premium puzzle. To address these claims we need to differentiate between two different interpretations of the term “equity premium”. One is the \textit{ex-post} or realized equity premium. This is the actual, \textit{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 is what Edward Prescott and I addressed in our paper. However, there is a related concept -- the \textit{ex-ante} equity premium. This is a forward-looking measure of the premium, that is, the equity premium that is \textit{expected} to prevail in the future or the conditional equity premium given the current state of the economy. To elaborate, after a bull market, when stock valuations are high relative to fundamentals the ex-ante equity premium is likely to be low. However, it is precisely in these times, when the market has risen sharply, that the ex-post, or the realized premium is high. Conversely, after a major downward correction, the ex-ante (expected) premium is likely to be high while the realized premium will be low. This should not come as a surprise since 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 this depends on the planning horizon. The equity premium that Edward Prescott and I documented is for very long investment horizons. It has little to do with what the premium is going to be over 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 depends on the time horizon over which it is measured.
There have even been periods when it has been negative, as illustrated in the figures below.

Source: Ibbotson 2001

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

The data used to document the equity premium over the past hundred years is as good an economic data set as we have and a hundred years is long series when it comes to economic data. Before we dismiss the premium, not only do we need to understand the observed phenomena but we also need a plausible explanation as to why the future is likely to be any different from the past. In the absence of this, and based on what we currently know, we can make the following claim: over the long horizon the equity premium is likely to be similar to what it has been in the past and the returns to investment in equity will continue to substantially dominate that in bills for investors with a long planning horizon.
References


