THE EQUITY PREMIUM IN RETROSPECT

Rajnish Mehra*
University of California, Santa Barbara and NBER
and
Edward C. Prescott*
University of Minnesota and Federal Reserve Bank of Minneapolis

Forthcoming in the *Handbook of the Economics of Finance*
Edited by G.M. Constantinides, M. Harris and R. Stulz
North Holland, Amsterdam

Current draft: Feb 2003

We thank George Constantinides, John Donaldson, Ellen R. McGrattan and Mark Rubinstein for helpful discussions. Mehra acknowledges financial support from the Academic Senate of the University of California. Prescott acknowledges financial support from the National Science Foundation.

More than two decades ago, we demonstrated that the equity premium (the return earned by a risky security in excess of that earned by a relatively risk-free T-bill), was an order of magnitude greater than could be rationalized in the context of the standard neoclassical paradigms of financial economics *as a premium for bearing risk*. We dubbed this historical regularity 'the equity premium puzzle.'(Mehra and Prescott(1985)). Our challenge to the profession has spawned a plethora of research efforts to explain it away.

In this paper, we take a retrospective look at the puzzle, critically examine the data sources used to document the puzzle, attempt to clearly explain it and evaluate the various attempts to solve it. The paper is organized into four parts. Part 1 documents the historical equity premium in the United States and in selected countries with significant capital markets in terms of market value and comments on the data sources. Part 2 examines the question, 'Is the equity premium due to a premium for bearing non-diversifiable risk?' Part 3 examines the related question, 'Is the equity premium due to borrowing constraints, a liquidity premium or taxes?' Finally, part 4 examines the equity premium expected to prevail in the future.

We conclude that research to date suggests that the answer to the first question is 'no', unless one is willing to accept that individuals are implausibly risk averse. In answer to the second question McGratten and Prescott (2001) found that, most likely, the high equity premium observed in the postwar period *was* indeed the result of a combination of the factors that included borrowing constraints and taxes.

1.1 Facts

Any discussion of the equity premium over time confronts the question of which average returns are more useful in summarizing historical information: arithmetic or geometric? It is well known that the arithmetic average return exceeds the geometric average return and that if the returns are log-normally distributed, the difference between the two is one-half the variance of the returns. Since the annual standard deviation of the equity premium is about 20 percent, this can result in a difference of about 2 percent between the two measures, which is non - trivial since the phenomena under consideration has an arithmetic mean of between 2 and 8 percent. In Mehra and Prescott (1985), we reported arithmetic averages, since the best available evidence indicated that stock returns were uncorrelated over time. When this is the case, the expected future value of a $1 investment is obtained by compounding the arithmetic average of the sample return, which is the correct statistic to report if one is interested in the *mean* value of the investment.[1] If, however, the objective is to obtain the *median* future value of the investment, then the initial investment should be compounded at the geometric sample average. When returns are serially correlated, then the arithmetic average[2] can lead to misleading estimates and thus the geometric average may be the more appropriate statistic to use. In this paper, as in our 1985 paper, we report arithmetic averages. However, in instances where we cite the results of research when arithmetic averages are not available, we clearly indicate this.[3]

1.2 Data Sources

A second crucial consideration in a discussion of the historical equity premium has to do with the reliability of early data sources. The data documenting the historical equity premium in the United States can be subdivided into three distinct sub-periods, 1802–1871, 1871–1926 and

[1] We present a simple proof in appendix A.
[2] The point is well illustrated by the textbook example where an initial investment of $100 is worth $200 after one year and $100 after two years. The arithmetic average return is 25% whereas the geometric average return is 0%. The latter coincides with the true return.
[3] In this case an approximate estimate of the arithmetic average return can be obtained by adding one-half the variance of the returns to the geometric average.

1926 – present. The quality of the data is very different for each subperiod. Data on stock prices for the nineteenth century is patchy, often necessarily introducing an element of arbitrariness to compensate for its incompleteness.

Subperiod 1802-1871

Equity Return Data

We find that the equity return data prior to 1871 is not particularly reliable. To the best of our knowledge, the stock return data used by all researchers for the period 1802–1871 is due to Schwert (1990), who gives an excellent account of the construction and composition of early stock market indexes. Schwert (1990) constructs a "spliced" index for the period 1802–1987; his index for the period 1802–1862 is based on the work of Smith and Cole (1935), who constructed a number of early stock indexes. For the period 1802–1820, their index was constructed from an equally weighted portfolio of seven bank stocks, while another index for 1815–1845 was composed of six bank stocks and one insurance stock. For the period 1834–1862 the index consisted of an equally weighted portfolio of (at most) 27 railroad stocks.[4] They used one price quote, per stock, per month, from local newspapers. The prices used were the average of the bid and ask prices, rather than transaction prices, and their computation of returns ignores dividends. For the period 1863–1871, Schwert uses data from Macaulay (1938), who constructed a value-weighted index using a portfolio of about 25 North-east and mid-Atlantic railroad stocks;[5] this index also excludes dividends. Needless to say, it is difficult to assess how well this data proxies the 'market,' since undoubtedly there were other industry sectors that were not reflected in the index.

[4] "They chose stocks in hindsight … the sample selection bias caused by including only stocks that survived and were actively quoted for the whole period is obvious." (Schwert (1990))

[5] "It is unclear what sources Macaulay used to collect individual stock prices but he included all railroads with actively traded stocks." Ibid

Return on a Risk-free Security

Since there were no Treasury bills at the time, researchers have used the data set constructed by Siegel (1998) for this period, using highly rated securities with an adjustment for the default premium. It is interesting to observe, as mentioned earlier, that based on this data set the equity premium for the period 1802–1862 was *zero.* We conjecture that this may be due to the fact that since most financing in the first half of the nineteenth century was done through debt, the distinction between debt and equity securities was not very clear-cut.[6]

Sub-period 1871–1926

Equity Return Data

Shiller (1989) is the definitive source for the equity return data for this period. His data is based on the work of Cowles (1939), which covers the period 1871–1938. Cowles used a value-weighted portfolio for his index, which consisted of 12 stocks[7] in 1871 and ended with 351 in 1938. He included all stocks listed on the New York Stock Exchange, whose prices were reported in the *Commercial and Financial Chronicle*. From 1918 onward he used the Standard and Poor's (S&P) industrial portfolios. Cowles reported dividends, so that, unlike the earlier indexes for the period 1802–1871, a total return calculation was possible.

Return on a Risk Free Security

There is no definitive source for the short-term risk-free rate in the period before 1920, when Treasury certificates were first issued. In our 1985 study, we used short-term commercial

[6] The first actively traded stock was floated in the U.S in 1791 and by 1801 there were over 300 corporations, although less than 10 were actively traded. (Siegel (1998)).
[7] It was only from Feb. 16, 1885, that Dow Jones began reporting an index, initially composed of 12 stocks. The S&P index dates back to 1928, though for the period 1928–1957 it consisted of 90 stocks. The S&P 500 debuted in March 1957.

paper as a proxy for a riskless short-term security prior to 1920 and Treasury certificates from 1920–1930. Our data prior to 1920, was taken from Homer (1963). Most researchers have either used our data set or Siegel's.

Sub-period 1926–present

Equity Return Data

This period is the "Golden Age" in regards to accurate financial data. The NYSE database at the Center for Research in Security Prices (CRSP) was initiated in 1926 and provides researchers with high quality equity return data. The Ibbotson Associates Yearbooks[8] are also a very useful compendium of post–1926 financial data.

Return on a Risk-free Security

Since the advent of Treasury bills in 1931, short maturity bills have been an excellent proxy for a "real" risk-free security since the innovation in inflation is orthogonal to the path of real GNP growth.[9] Of course, with the advent of Treasury Inflation Protected Securities (TIPS) on January 29, 1997, the return on these securities *is* the real risk-free rate.

1.3 Estimates of the Equity Premium

Historical data provides us with a wealth of evidence documenting that for over a century, stock returns have been considerably higher than those for Treasury-bills. This is illustrated in Table 1, which reports the unconditional estimates[10] for the US equity premium based on the

---

[8] Ibbotson Associates. "Stocks, Bonds, Bills and Inflation." 2000 Yearbook. Chicago. Ibbotson Associates. 2001.

[9] See Litterman (1980) who also found that that in post war data the innovation in inflation had a standard deviation of one half of one percent.

[10] To obtain unconditional estimates we use the entire data set to form our estimate. The Mehra-Prescott data set covers the longest time period for which ***both*** consumption and stock return data is available. The former is necessary to test the implication of consumption based asset pricing models.

various data sets used in the literature, going back to 1802. The average annual real return, (the inflation adjusted return) on the U.S. stock market over the last 110 years has been about 8.06 percent. Over the same period, the return on a relatively riskless security was a paltry 1.14 percent. The difference between these two returns, the "equity premium," was 6.92 percent.

Furthermore, this pattern of excess returns to equity holdings is not unique to the U.S. but is observed in every country with a significant capital market. The U.S. together with the U.K., Japan, Germany and France accounts for more than 85 percent of the capitalized global equity value.

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 and Japan. Table 2 illustrates the equity premium in the post war period for these countries.

Table 1

U.S. Equity Premium Using Different Data Sets

Data Set	% real return on a market index	% real return on a relatively riskless security	% equity premium
	Mean	Mean	Mean
1802-1998 (Siegel)	7.0	2.9	4.1
1871-199 (Shiller)	6.99	1.74	5.75
1889-2000 (Mehra-Prescott)	8.06	1.14	6.92
1926-2000 (Ibbotson)	8.8	0.4	8.4

Table 2 Equity Premium in Different Countries			
Country	% real return on a market index	% real return on a relatively riskless security	% equity premium
	Mean	Mean	Mean
UK (1947-1999)	5.7	1.1	4.6
Japan (1970-1999)	4.7	1.4	3.3
Germany (1978-1997)	9.8	3.2	6.6
France (1973-1998)	9.0	2.7	6.3
Source: U.K from Siegel (1998), the rest are from Campbell (2001)			

The dramatic investment implications of this differential rate of return can be seen in Table 3, which maps the capital appreciation of $1 invested in different assets from 1802 to 1997 and from 1926 to 2000.

Table 3 Terminal value of $1 invested in Stocks and Bonds				
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
Source: Ibbotson (2001) and Siegel (1998)				

As Table 3 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. The corresponding values for the 75-year period, 1926–2000, 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.

In Table 4 we report the premium for some interesting sub-periods: 1889–1933, when the United States was on a gold standard; 1933–2000, when it was off the gold standard; and 1946–2000, the postwar period. Table 5 presents 30 year moving averages, similar to those reported by the US meteorological service to document 'normal' temperature.

Table 4

Equity Premium in Different Sub-Periods

Time Period	% real return on a market index	% real return on a relatively riskless security	% equity premium
	Mean	Mean	Mean
1889–1933	7.01	3.09	3.92
1934–2000	8.76	-0.17	8.93
1946–2000	9.03	0.68	8.36

Source: Mehra and Prescott (1985). Updated by the authors.

Table 5

Equity Premium: 30 Year Moving Averages

Time Period	% real return on a market index	% real return on a relatively riskless security	% equity premium
	Mean	Mean	Mean
1900–1950	6.51	2.01	4.50
1951–2000	8.98	1.41	7.58

Source: Mehra and Prescott (1985). Updated by the authors

Although the premium has been increasing over time, this is largely due to the diminishing return on the riskless asset, rather than a dramatic increase in the return on equity, which has been relatively constant. The low premium in the nineteenth centaury is largely due to the fact that the equity premium for the period 1802–1861 was zero.[11] If we exclude this period, we find that difference in the premium in the second half of the nineteenth century relative to average values in the twentieth century is less striking.

We find a dramatic change in the equity premium in the post 1933 period – the premium rose from 3.92 percent to 8.93 percent, an increase of more than 125 percent. Since 1933 marked the end of the period when the US was on the gold standard, this break can be seen as the change in the equity premium after the implementation of the new policy.

1.4 Variation in the Equity Premium over Time

The equity premium has varied considerably over time, as illustrated in Figures 1 and 2, below. Furthermore, the variation depends on the time horizon over which it is measured. There

[11] See the earlier discussion on data.

have even been periods when it has been negative.

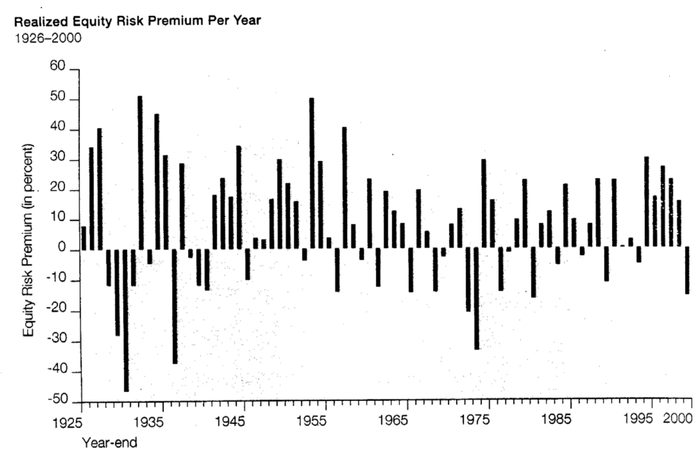


Figure 1

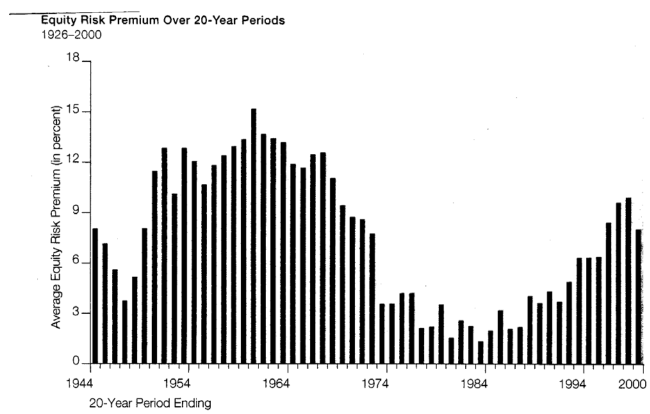


Figure 2

Source: Ibbotson 2001

The low frequency variation has been *counter- cyclical*. This is shown in Figure 3 where we have plotted stock market value as a share of national income[12] and the mean equity premium averaged over certain time periods. We have divided the time period from 1929 to 2000 into sub-periods, where the ratio market value of equity to national income was greater than 1 and where it was less than 1. Historically, as the figure illustrates, subsequent to periods when this ratio was

[12] In Mehra (1998) it is argued that the variation in this ratio is difficult to rationalize in the standard neoclassical framework since over the same period after tax cash flows to equity as a share of national income are fairly constant. Here we do not address

high the realized equity premium was low. A similar results holds when stock valuations are low relative to national income. In this case the subsequent equity premium is high.

Since After Tax Corporate Profits as a share of National Income are fairly constant over

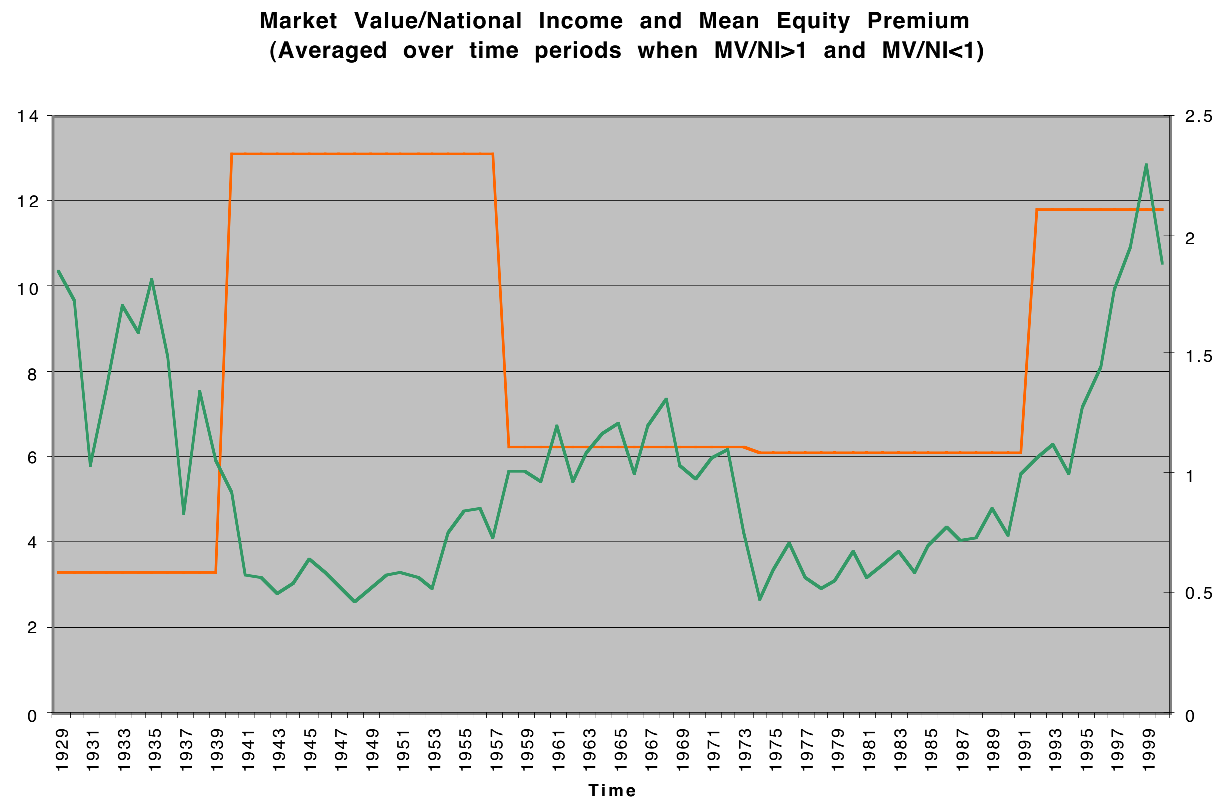


Figure 3

time, this translates into the observation that the realized equity premium was low subsequent to periods when the Price/ Earnings ratio is high and vice versa. This is the basis for the returns predictability literature in Finance. (Campbell and Shiller (1988) and Fama and French (1988)).

---

this issue and simply utilize the fact that this *ratio has varied considerably over time.*

In Figure 4we have plotted stock market value as a share of national income and the *sub-sequent* three-year mean equity premium. This provides further conformation that historically,

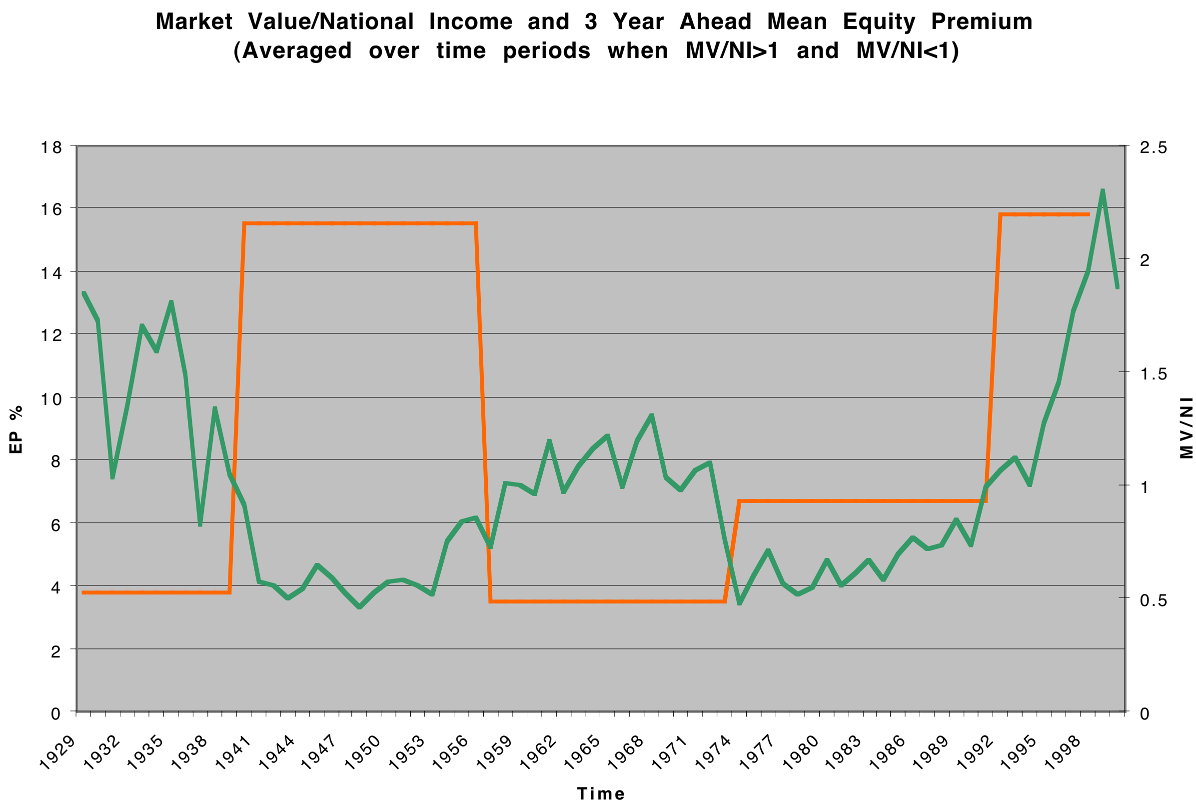


Figure 4

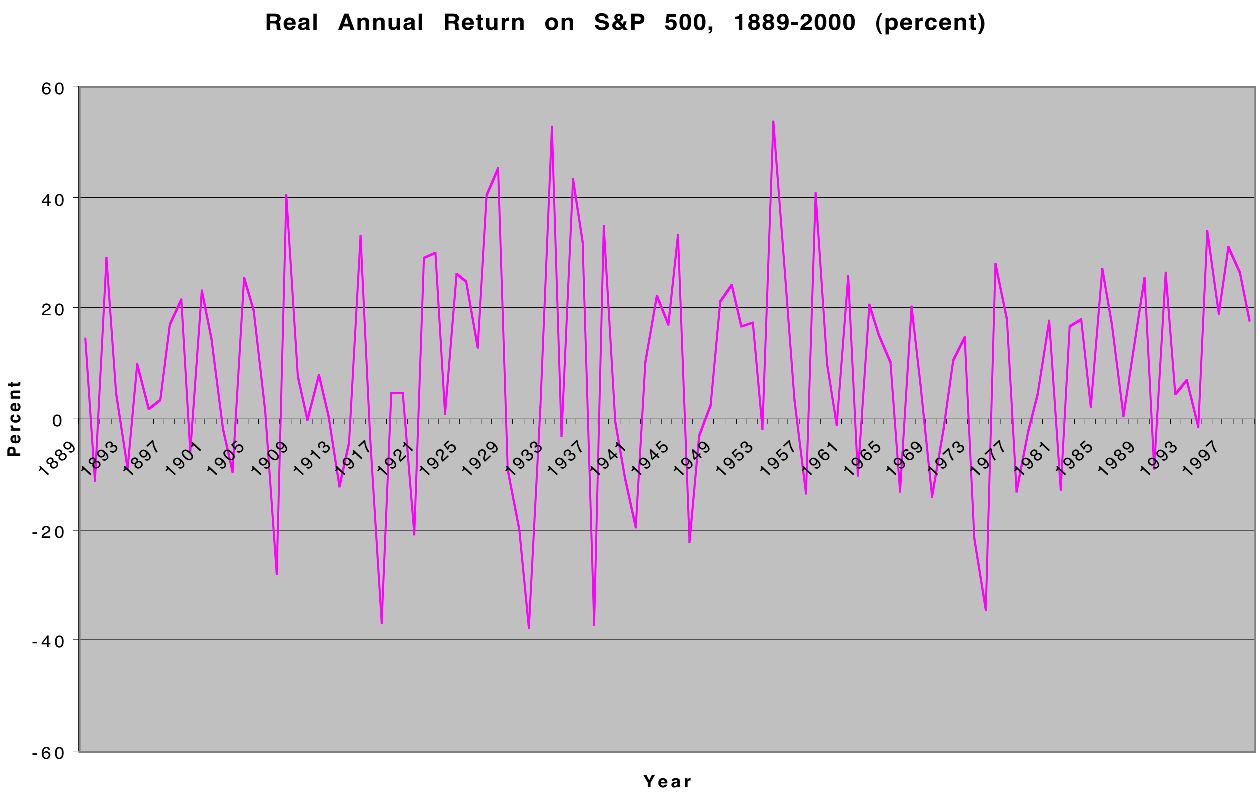
periods of relatively high market valuation have been followed by periods when the equity premium was relatively low.

2. Is The Equity Premium Due To A Premium For Bearing Non-diversifiable Risk?

In this section, we examine various models that attempt to explain the historical equity premium. We start with a model with standard (CRRA) preferences, then examine models incorporating alternative preference structures, idiosyncratic and uninsurable income risk, and models incorporating a disaster state and survivorship bias.

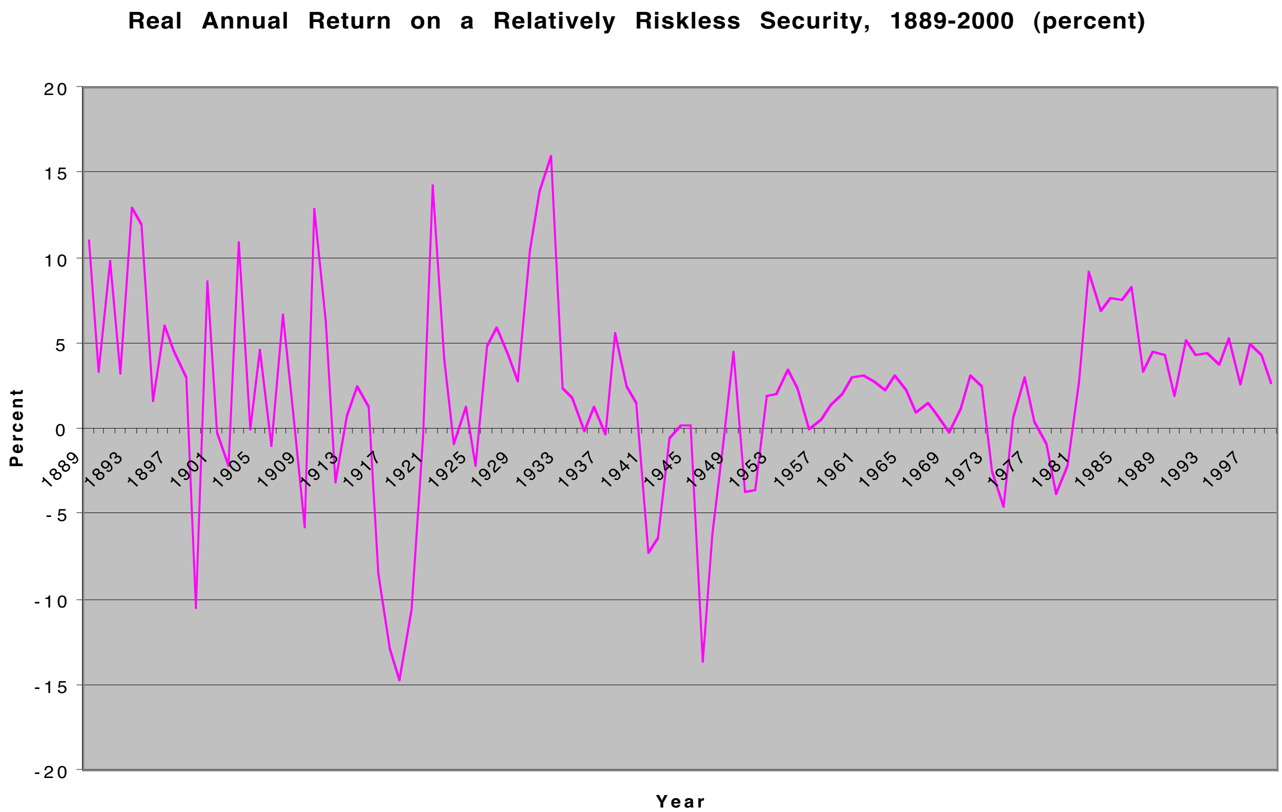
Why have stocks been such an attractive investment relative to bonds? Why has the rate of return on stocks been higher than on relatively risk-free assets? One intuitive answer is that since stocks are 'riskier' than bonds, investors require a larger premium for bearing this additional risk; and indeed, the standard deviation of the returns to stocks (about 20 percent per annum historically) is larger than that of the returns to T-bills (about 4 percent per annum), so, obviously they are considerably more risky than bills! But are they?

Figures 5 and 6 below illustrate the variability of the annual real rate of return on the S&P 500 index and a relatively risk-free security over the period 1889–2000.Of course, the index did not consist of 500 stocks for the entire period.



Source: Mehra and Prescott (1985). Data updated by the authors.

Figure 5



Source: Mehra and Prescott (1985). Data updated by the authors.

Figure 6

To enhance and deepen our understanding of the risk-return trade-off in the pricing of financial assets, we take a detour into modern asset pricing theory and look at why different assets yield different rates of return. The deus 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.

The operative emphasis here is the *incremental loss or gain* of utility of 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. As a consequence, assets that pay off when times are good and consumption levels are high – when the marginal utility of consumption is low – are less desirable than those that pay off an equivalent amount when times are bad and additional consumption is more highly valued. Hence consumption in period *t* has a different price if times are good than if times are bad.

Let us illustrate this principle in the context of the standard, popular paradigm, the Capital Asset Pricing Model (CAPM). The model postulates a linear relationship between an asset's 'beta,' a measure of systematic risk, and its expected return. Thus, high-beta stocks yield a high expected rate of return. That is 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 – when times are good and consumption is plentiful; it provides less incremental utility than a security that pays off when consumption is low, is less valuable and consequently sells for less. Thus higher beta 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, the lower beta 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 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, despite of their very low rates of return).

To return to the original question: are stocks that much riskier than T-bills so as to justify a six percentage differential in their rates of return?

What came as a surprise to many economists and researchers in finance was the conclusion of a paper by Mehra and Prescott, written 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 one percent return premium over bills. Since, for as long as we had reliable data (about 100 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)).

2.1 Standard Preferences

The neoclassical growth model and its stochastic variants are a central construct in contemporary finance, public finance, and business cycle theory. It has been used extensively by, among others, Abel et al. (1989), Auerbach and Kotlikoff (1987), Barro and Becker (1988), Brock (1979), Cox, Ingersoll and Ross (1985), Donaldson and Mehra (1984), Lucas (1978), Kydland and Prescott (1982), and Merton (1971). In fact, much of our economic intuition is derived from this model class. A key idea of this framework is that consumption today and consumption in some future period are treated as different goods. Relative prices of these different goods are equal to people's willingness to substitute between these goods and businesses' ability to transform these goods into each other.

The model has had some remarkable successes when confronted with empirical data, particularly in the stream of macroeconomic research referred to as Real Business Cycle Theory, where researchers have found that it easily replicates the essential macroeconomic features of the business cycle. See, in particular, Kydland and Prescott (1982). Unfortunately, when confronted with financial market data on stock returns, tests of these models have led, without exception, to their rejection. Perhaps the most striking of these rejections is contained in our 1985 paper.

To illustrate this we employ a variation of Lucas' (1978) pure exchange model. Since per capita consumption has grown over time, we assume that the growth rate of the endowment follows a Markov process. This is in contrast to the assumption in Lucas' model that the endowment level follows a Markov process. Our assumption, which requires an extension of competitive equilibrium theory[13], enables us to capture the non-stationarity in the consumption series associated with the large increase in per capita consumption that occurred over the last century.

We consider a frictionless economy that has a single representative 'stand-in' household. This unit orders its preferences over random consumption paths by

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

where c_t is the per capita consumption and the parameter β is the subjective time discount factor, which describes how impatient households are to consume. If β is small, people are highly impatient, with a strong preference for consumption now versus consumption in the future. As modeled, these households live forever, which implicitly means that the utility of parents depends on the utility of their children. In the real world, this is true for some people and not for others. However, economies with both types of people—those who care about their children's utility and those who do not—have essentially the same implications for asset prices and returns.[14]

Thus, we use this simple abstraction to build quantitative economic intuition about what the returns on equity and debt should be. $E_0\{\cdot\}$ is the expectation operator conditional upon information available at time zero, (which denotes the present time) and U: $\mathrm{R}_+ \rightarrow \mathrm{R}$ is the increasing, continuously differentiable concave utility function. We further restrict the utility function to be of the constant relative risk aversion (CRRA) class

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

where the parameter α measures the curvature of the utility function. When α = 1, the utility function is defined to be logarithmic, which is the limit of the above representation as α approaches 1. The feature that makes this the "preference function of choice" in much of the literature in Growth and Real Business Cycle Theory is that it is scale invariant. This means that a household is more likely to accept a gamble if both its wealth and the gamble amount are scaled by a positive factor. Hence, although the level of aggregate variables such as capital stock have increased over time, the resulting equilibrium return process is stationary. A second attractive feature is that it is one of only two preference functions that allows 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.

---

[13] This extension is developed Mehra (1988).
[14] See Constantinides, Donaldson and Mehra (2002).

There is no fundamental economic reason why this must be so. We will revisit this issue in Section 3, where we examine preference structures that do not impose this restriction.[15]

We assume there is one productive unit which produces output y_t in period t which is the period dividend. There is one equity share with price p_t that is competitively traded; it is a claim to 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 in the 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 in the 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. Hence the fundamental relation that prices assets is $p_t U'(c_t) = \beta E_t\{(p_{t+1} + y_{t+1})U'(c_{t+1})\}$. Versions of this expression can be found in Rubinstein (1976), Lucas (1978), Breeden (1979), and Prescott and Mehra (1980), among others. Excellent textbook treatments can be found in Cochrane (2001), Danthine and Donaldson (2001), Duffie (2001), and LeRoy and Werner (2001).

We use it to price both stocks and risk-less one period bonds.

For equity we have

$$1 = \beta E_t\left\{\frac{U'(c_{t+1})}{U'(c_t)} R_{e,t+1}\right\} \tag{3}$$

where

$$R_{e,t+1} = \frac{p_{t+1} + y_{t+1}}{p_t} \tag{4}$$

[15] Epstein and Zin (1991) and Weil (1989).

and for the risk-less one period bonds the relevant expression is

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

Where the gross rate of return on the riskless asset is by definition

$$R_{f,t+1}=\frac{1}{q_t} \tag{6}$$

with q_t being the price of the bond. Since $U(c)$ is assumed to be increasing, we can rewrite (3) as

$$1=\beta E_t\left\{M_{t+1}R_{e,t+1}\right\} \tag{7}$$

where M_{t+1} is a strictly positive stochastic discount factor. This guarantees that the economy will be arbitrage free and the law of one-price holds. A little algebra shows that

$$E_t(R_{e,t+1})=R_{f,t+1}+Cov_t\left\{\frac{-U'(c_{t+1}),R_{e,t+1}}{E_t(U'(c_{t+1}))}\right\} \tag{8}$$

The equity premium $E_t(R_{e,t+1})-R_{f,t+1}$ thus can 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 co-vary positively with consumption – that is, they payoff in states when consumption is high and marginal utility is low – command a high premium since these assets "destabilize" consumption.

The question we need to address is the following: is the magnitude of the covariance between the marginal utility of consumption large enough to justify the observed 6 percent equity premium in U.S. equity markets? If not, how much of historic equity premium is a premium for bearing non-diversifiable aggregate risk.

To address this issue, we present a variation on the framework used in our original paper on the equity premium. An advantage of our original approach was that we could easily test the sensitivity of our results to changes in distributional assumptions[16]. We found that our results were essentially unchanged for very different consumption processes, provided that the mean and variances of growth rates equaled the historically observed values and the coefficient of relative risk aversion was less than ten[17]. Using this insight on the robustness of the results to distributional assumptions from our earlier analysis we consider the case where the growth rate of consumption $x_{t+1} \equiv \dfrac{c_{t+1}}{c_t}$ is i.i.d and lognormal. We do this to facilitate exposition and because this results in closed form solutions[18].

As a consequence, the gross return on equity $R_{e,t}$ (defined above) is i.i.d, and log-normal. Substituting $U'(c_t) = c_t^{-\alpha}$ in the fundamental pricing relation and noting that in this exchange economy the equilibrium consumption process is $\{y_t\}$

$$p_t = \beta E_t \left\{ (p_{t+1} + y_{t+1}) U'(c_{t+1}) / U'(c_t) \right\} \tag{9}$$

we get

$$p_t = \beta E_t \left\{ (p_{t+1} + y_{t+1}) x_{t+1}^{-\alpha} \right\} \tag{10}$$

As p_t is homogeneous of degree one in y_t we can represent it as

$$p_t = w y_t$$

and hence $R_{e,t+1}$ can be expressed as

$$R_{e,t+1} = \frac{(w+1)}{w} \cdot \frac{y_{t+1}}{y_t} = \frac{w+1}{w} \cdot x_{t+1} \tag{11}$$

---

[16] In contrast to our approach, which is in the applied general equilibrium tradition, there is another tradition of testing Euler equations (such as equation 9) and rejecting them. Hansen and Singleton (1982) and Grossman and Shiller (1981) exemplify this approach.

[17] See Mehra and Prescott (1985) pages 156-57. The original framework also allowed us to address the issue of leverage.

[18] The exposition below is based on Abel (1988). Our original analysis is presented in Appendix B

It is easily shown[19] that

$$w = \frac{\beta E_t\{x_{t+1}^{1-\alpha}\}}{1-\beta E_t\{x_{t+1}^{1-\alpha}\}} \tag{12}$$

hence
$$E_t\{R_{e,t+1}\} = \frac{E_t\{x_{t+1}\}}{\beta E_t\{x_{t+1}^{1-\alpha}\}} \tag{13}$$

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

$$R_{f,t+1} = \frac{1}{\beta}\,\frac{1}{E_t\{x_{t+1}^{-\alpha}\}} \tag{14}$$

Since we have assumed the growth rate of consumption and dividends to be log normally distributed,

$$E_t\{R_{e,t+1}\} = \frac{e^{\mu_x + 1/2\sigma_x^2}}{\beta e^{(1-\alpha)\mu_x + 1/2(1-\alpha)^2\sigma_x^2}} \tag{15}$$

and
$$\ln E_t\{R_{e,t+1}\} = -\ln\beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2 + \alpha\sigma_x^2 \tag{16}$$

where $\mu_x = E(\ln x)$, $\sigma_x^2 = Var(\ln x)$ and $\ln x$ is the *continuously compounded* growth rate of consumption.

Similarly
$$R_f = \frac{1}{\beta e^{-\alpha\mu_x + 1/2\alpha^2\sigma_x^2}} \tag{17}$$

and
$$\ln R_f = -\ln\beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2 \tag{18}$$

$\therefore$
$$\ln E\{R_e\} - \ln R_f = \alpha\sigma_x^2 \tag{19}$$

From (11) it also follows that
$$\ln E\{R_e\} - \ln R_f = \alpha\sigma_{x,R_e} \tag{20}$$

where
$$\sigma_{x,R_e} = Cov(\ln x, \ln R_e) \tag{21}$$

[19] See Appendix A in Mehra (2003)

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. From equation 19, it is also the product of the coefficient of relative risk aversion and the variance of the growth rate of consumption. As we see below, this variance σ_x^2 is 0.00125, so unless the coefficient of risk aversion α, is large, a high equity premium is impossible. The growth rate of consumption just does not vary enough!

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

Mean 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 the growth rate of consumption $\sigma\{x\}$	0.036
Mean equity premium $E\{R_e\}- R_f$	0.0618

In our calibration, we are guided by the tenet that model parameters should meet the criteria of 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. A number of these studies are documented in Mehra and Prescott (1985). We can then pose a question: if we set the 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 above?

Using the expressions derived earlier we have

$$\ln R_f = -\ln\beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2 = 0.120$$

or $$R_f = 1.127$$

that is, a risk free rate of 12.7%!

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

$$= 0.132$$

we have $$E\{R_e\} = 1.141$$

or a return on equity of 14.1%. This implies an equity risk premium of 1.4%, far lower than the 6.18% historically observed equity premium. In this calculation we have been very liberal in choosing the values for α and β. Most studies indicate a value for α that is close to 3. If we pick a lower value for β, the riskfree rate will be even higher and the premium lower. So the 1.4% value represents the maximum equity risk premium that can be obtained in this class of models given the constraints on α and β. Since the observed equity premium is over 6%, we have a puzzle on our hands that risk considerations alone cannot account for.

Philippe Weil (1989) has dubbed the high risk-free rate obtained above "the risk-free rate puzzle." The short-term real rate in the United States averages less than 1 percent, while the high value of α required to generate the observed equity premium results in an unacceptably high risk-free rate. The risk-free rate as shown in equation (18) can be decomposed into three components:

$$\ln R_f = -\ln \beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2$$

The first term, $-\ln \beta$, is a time preference or impatience term. When $\beta < 1$ it reflects the fact that agents prefer early consumption to later consumption. Thus in a world of perfect certainty and no growth in consumption, the unique interest rate in the economy will be $R_f = 1/\beta$.

The second term, $\alpha\mu_x$, arises because of growth in consumption. If consumption is likely to be higher in the future, agents with concave utility would like to borrow against future consumption in order to smooth their lifetime consumption. The higher the curvature of the utility function and the larger the growth rate of consumption, the greater the desire to smooth consumption. In equilibrium this will lead to a higher interest rate since agents in the aggregate cannot simultaneously increase their current consumption.

The third term $1/2\alpha^2\sigma_x^2$ arises due to a demand for precautionary saving. In a world of uncertainty, agents would like to hedge against future unfavorable consumption realizations by building " buffer stocks" of the consumption good. Hence, in equilibrium, the interest rate must fall to counter this enhanced demand for savings.

Figure 7, below, plots $\ln R_f = -\ln \beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2$ calibrated to the U.S. historical values with $\mu_x = 0.0175$ and $\sigma_x^2 = 0.00123$ for various values of β. It shows that the precautionary savings effect is negligible for reasonable values of α. ($1<\alpha<5$)

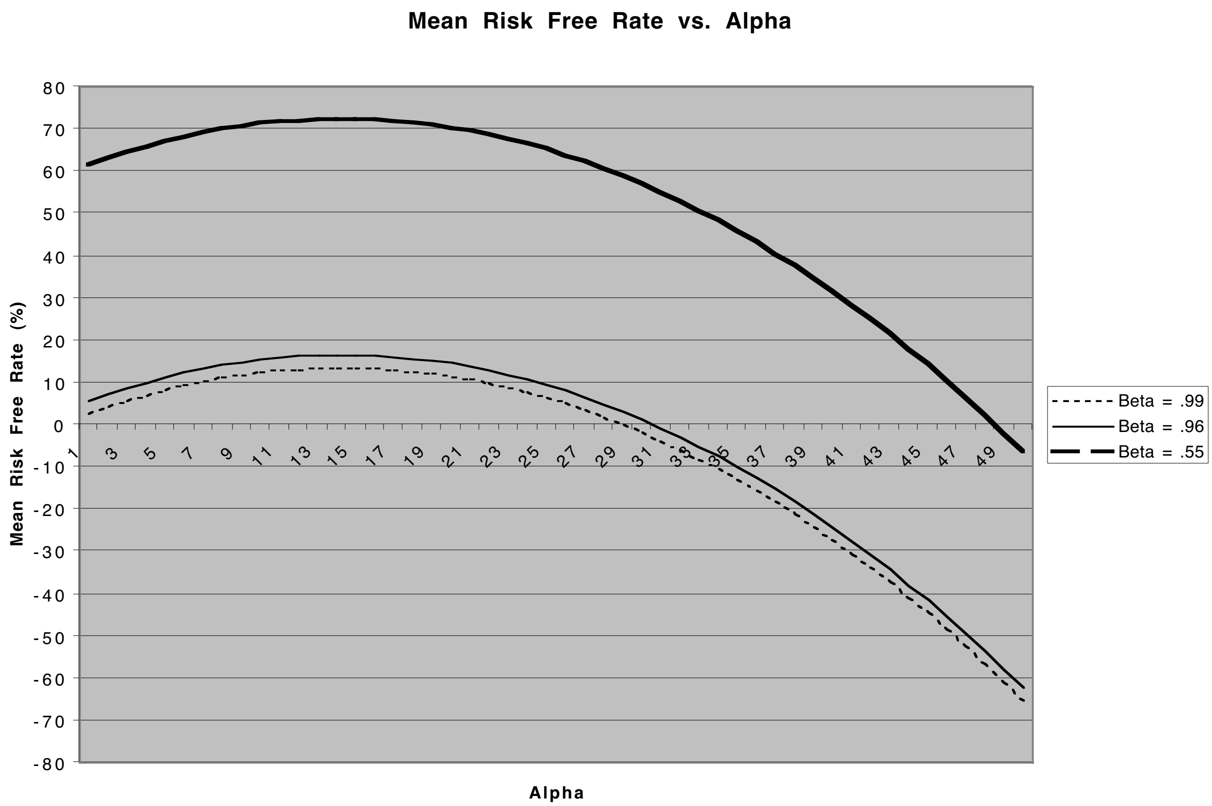


Figure 7

For $\alpha = 3$ and $\beta = 0.99$ R_f = 1.65, which implies a risk-free rate of 6.5 percent – much higher than the historical mean rate of 0.8 percent. The economic intuition is straightforward – with consumption growing at 1.8 percent a year with a standard deviation of 3.6 percent, agents with isoelastic preferences have a sufficiently strong desire to borrow to smooth consumption that it takes a high interest rate to induce them not to do so.

The late Fischer Black[20] proposed that α =55 would solve the puzzle. Indeed it can be shown that the 1889–1978 U.S. experience reported above can be reconciled with α = 48 and β=0.55.

To see this, observe that since $\sigma_x^2 = \ln\left[1+\frac{\text{var}(x)}{[E(x)]^2}\right]$ = 0.00125

and $\mu_x = \ln E(x) - 1/2\,\sigma_x^2 = 0.0172$

this implies $$\alpha = \frac{\ln E(R) - \ln R_F}{\alpha_x^2}$$

$$= 47.6$$

Since $\ln \beta = -\ln R_F + \alpha\mu_x - 1/2\alpha^2 \sigma_x^2$

$$= -0.60$$

this implies $\beta = 0.55$.

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 unacceptable risk-free rate as shown in Figure 6. An alternate approach is to experiment with negative time preferences; however there seems to be no empirical evidence that agents do have such preferences.[21]

---

[20] Private communication 1981.
[21] In a model with growth equilibrium can exist with β >1. See Mehra (1988) for the restrictions on the parameters α and β for equilibrium to exist.

Figure 7 shows that for extremely high α the precautionary savings term dominates and results in a "low" risk-free rate.[22] 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 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 was large. In section 3 we show how alternative preference structures can help resolve the risk free rate puzzle.

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

$$E_t(R_{e,t+1}) = R_{f,t+1} + Cov_t\left\{\frac{M_{t+1}, R_{e,t+1}}{E_t(M_{t+1})}\right\} \tag{22}$$

This expression also holds unconditionally so that

$$E(R_{e,t+1}) = R_{f,t+1} + \sigma(M_{t+1})\sigma(R_{e,t+1})\rho_{R,M} / E\ (M_{t+1}) \tag{23}$$

or

$$E(R_{e,t+1}) - R_{f,t+1} / \sigma(R_{e,t+1}) = \sigma(M_{t+1})\rho_{R,M} / E(M_{t+1}) \tag{24}$$

and since $-1 \le \rho_{R,M} \le 1$

$$\left|E(R_{e,t+1}) - R_{f,t+1} / \sigma(R_{e,t+1})\right| \le \sigma(M_{t+1}) / E(M_{t+1}) \tag{25}$$

This inequality is referred to as the Hansen-Jagannathan lower bound on the pricing kernel.

For the U.S. economy, the Sharpe Ratio, $E(R_{e,t+1}) - R_{f,t+1} / \sigma(R_{e,t+1})$, can be calculated to be 0.37. Since $E(M_{t+1})$ 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 α =2. This im-

[22]Kandel and Stambaugh (1991) have suggested this approach.

plies that the lower bound on the standard deviation for the pricing kernel must be close to 0.3 if the Hansen-Jagannathan bound is to be satisfied. However, when this is calculated in the Mehra-Prescott framework, we obtain an estimate for $\sigma(M_{t+1})$=0.002, which is off by more than an order of magnitude.

We would like 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, 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. Several generalizations of key features of the Mehra and Prescott (1985) model have been proposed to better reconcile observations with theory. These include alternative assumptions on preferences,[23] modified probability distributions to admit rare but disastrous events,[24]survival bias,[25] incomplete markets,[26] and market imperfec-

[23] For example, Abel (1990), Bansal and Yaron (2000), Benartzi and Thaler (1995), Boldrin, Christiano and Fisher (2001), Campbell and Cochrane (1999), Constantinides (1990), Epstein and Zin (1991), and Ferson and Constantinides (1991).
[24] See, Rietz (1988) and Mehra and Prescott (1988).
[25]Brown, Goetzmann and Ross (1995)
[26] For example, Bewley (1982), Brav, Constantinides and Geczy (2002), Constantinides and Duffie (1996), Heaton and Lucas (1997, 2000), Krebs (2000), Lucas (1994), Mankiw (1986), Mehra and Prescott (1985), Storesletten, Telmer and Yaron (2001), and Telmer (1993).

tions.[27] They also include attempts at modeling limited participation of consumers in the stock market,[28] and problems of temporal aggregation.[29] However, none have fully resolved the anomalies. We examine some of the research efforts to resolve the puzzle[30] below and in Section 3.

2.2 Estimating equity risk premium versus estimating risk aversion parameter.

Estimating or measuring the relative risk parameter using statistical tools is very different than estimating the equity risk premium. Mehra and Prescott (1985), as discussed above, use an extension of Lucas' (1978) asset pricing model to estimate how much of the historical difference in yields on treasury bills and corporate equity is a premium for bearing aggregate risk. Crucial to their analysis is their use of micro observations to restrict the value of the risk aversion parameter. *They did not estimate either the risk aversion parameter or the discount rate parameters.* Mehra and Prescott (1985) reject extreme risk aversion based upon observations on individual behavior. These observations include the small size of premia for jobs with uncertain income and the limited amount of insurance against idiosyncratic income risk. Another observation is that people with limited access to capital markets make investments in human capital that result in very uneven consumption over time.

[27] For example, Aiyagari and Gertler (1991), Alvarez and Jerman (2000), Bansal and Coleman (1996), Basak and Cuoco (1998), Constantinides, Donaldson and Mehra (2002), Danthine, Donaldson and Mehra (1992), Daniel and Marshall (1997), He and Modest (1995), Heaton and Lucas (1996), and Luttmer (1996), McGrattan and Prescott (2001), and Storesletten, Telmer and Yaron (1999).
[28] Attanasio, Banks and Tanner (2002), Brav, Constantinides and Geczy (2002), Brav and Geczy (1995), Mankiw and Zeldes (1991), and Vissing-Jorgensen (2002).
[29] Gabaix and Laibson (2001), Heaton (1995), and Lynch (1996).
[30] The reader is also referred to the excellent surveys by Narayana Kocherlakota (1996), John Cochrane (1997) and by John Campbell (1999,2001).

A sharp estimate for the magnitude of the risk aversion parameter comes from macroeconomics. The evidence is that the basic growth model, when restricted to be consistent with the growth facts, generates business cycle fluctuations if and only if this risk aversion parameter is near zero. (This corresponds to the log case in standard usage). The point is that the risk aversion parameter comes up in wide variety of observations at both the household and the aggregate level and is *not found to be large.*

For all values of the risk-aversion coefficient less than ten, which is an upper bound number for this parameter, Mehra and Prescott find that a premium for bearing aggregate risk accounts for little of the historic equity premium. This finding has stood the test of time.

Another tradition is to use consumption and stock market data to estimate the degree of relative risk aversion parameter and the discount factor parameter. This is what Grossman and Shiller report they did in their *American Economic Review Papers and Proceedings* article (1981, p. 226). Hansen and Singleton in a paper in which they develop "a method for estimating nonlinear rational expectations models directly from stochastic Euler equations" illustrate their methods by estimating the risk aversion parameter and the discount factor using stock dividend consumption prices (1981, p. 1269).

What the work of Grossman and Shiller (ibid) and Hansen and Singleton (ibid) establish is that using consumption and stock market data and assuming frictionless capital markets is a bad way to estimate the risk aversion and discount factor parameters. It is analogous to estimating the force of gravity near the earth's surface by dropping a feather from the top of the Leaning Tower of Pisa, under the assumption that friction is zero.

A tradition related to the statistical estimation is to statistically test whether the stochastic Euler equation arising from the stand-in household's inter-temporal optimization holds. Both

Grossman (1983) and Hansen-Singleton (1981) reject this relation. The fact that this relation is inconsistent with the U.S. time series data is no reason to conclude that the model economy used by Mehra and Prescott to estimation how much of the historical equity premium is a premium for bearing aggregate risk. Returning to the analogy from Physics, it would be silly to reject Newtonian mechanics as a useful tool for drawing scientific inference because the distance traveled by the feather did not satisfy $\frac{1}{2} g t^2$.

2.2 Alternative Preference Structures

Modifying the conventional time - and state - separable utility function

The analysis above shows that the isoelastic preferences used in Mehra and Prescott (1985) can only be made consistent with the observed equity premium if the coefficient of relative risk aversion is implausibly large. One 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. What this implies is that if an individual is averse to variation of consumption across different states *at a particular point of time* then he will be averse to *consumption variation over time*. There is no a priori reason that this must be so. Since, on average, consumption is growing over time, the agents in the Mehra and Prescott (1985) setup have little incentive to save. The demand for bonds is low and as a consequence the risk-free rate is counterfactually high. Epstein and Zin (1991) have presented a class of preferences that they term "Generalized Expected Utility" (GEU) which allows independent parameterization for the coefficient of risk aversion and the elasticity of intertemporal substitution.

In this class of preferences utility is recursively defined by

$$U_t = \{(1-\beta)c_t^{\rho} + \beta\{E_t(\tilde{U}_{t+1}^{\alpha})\}^{\frac{\rho}{\alpha}}\}^{\frac{1}{\rho}} \qquad (26)$$

where $1-\alpha$ is the coefficient of relative risk aversion and $\sigma = \frac{1}{1-\rho}$ the elasticity of intertemporal substitution. The usual iso-elastic preferences follow as a special case when $\rho = \alpha$. In the Epstein and Zin model, agents' wealth *W* evolves as $W_{t+1} = (W_t - c_t)(1 + R_{w,t+1})$ where $R_{w,t+1}$ is the return on *all* invested wealth and is potentially *unobservable*. To examine the asset pricing implications of this modification we examine the pricing kernel[31]

$$k_{t+1} = \beta^{\frac{\alpha}{\rho}} \left(\frac{c_{t+1}}{c_t} \right)^{\frac{\alpha(\rho-1)}{\rho}} (1 + R_{w,t+1})^{\frac{\alpha-\rho}{\rho}} \tag{27}$$

Thus the price p_t of an asset with payoff y_{t+1} at time t+1 is

$$p_t = E_t(k_{t+1} y_{t+1}) \tag{28}$$

In this framework the asset is priced *both* by its covariance with the growth rate of consumption (the first term in 27) *and* with the return on the wealth portfolio. This captures the pricing features of both the standard consumption CAPM and the traditional static CAPM. To see this, note that when $\alpha = \rho$, we get the consumption CAPM and with logarithmic preferences ($\alpha / \rho = 0$), the static CAPM.

Another feature of this class of models is that a high coefficient of risk aversion, 1-α, does not necessarily imply that agents will want to smooth consumption over time. However, the main difficulty in testing this alternative preference structure stems from the fact that the counterparts of equations (3) and (5) using GEU depend on variables that are *unobservable*, and this makes calibration tricky. One needs to make specific assumptions on the consumption process to obtain first order conditions in terms of observables. Epstein and Zin (1991) use the "market portfolio"

[31] Epstein and Zin (1991) use dynamic programming to calculate this. See their equations 8-13. Although the final result is correct although there appears to be errors in the intermediate steps.

as a proxy for the wealth portfolio and claim that their framework offers a solution to the equity premium puzzle. We feel that this proxy overstates the correlation between asset returns and the wealth portfolio and hence their claim.

This modification has the *potential* to resolve the risk free rate puzzle. We illustrate this below. Under the log-normality assumptions from section 2, and using the market portfolio as a stand in for the wealth portfolio we have

$$\ln R_f = -\ln \beta + \frac{\mu_x}{\sigma} - \frac{\alpha / \rho}{2\sigma^2}\sigma_x^2 + \frac{(\alpha / \rho) - 1}{2}\sigma_m^2 \qquad (29)$$

Here σ_m^2 is the variance of the return on the "market portfolio" of all invested wealth. Since $1-\alpha$ need not equal $1/\sigma$, we can have a large α without making σ small and hence obtain a reasonable risk free rate if one is prepared to assume a large σ. The problem with this is that there is independent evidence that the elasticity of intertemporal substitution is *small* (Campbell (2001)) hence this generality is not very useful when the model is accurately calibrated.

Habit Formation.

A second approach to modifying preferences was initiated by Constantinides (1990) by incorporating habit formation. This formulation assumes that utility is affected not only by current consumption but also by past consumption. It captures a fundamental feature of human behavior that repeated exposure to a stimulus diminishes the response to it. The literature distinguishes between two types of habit, "internal" and "external" and two modeling perspectives, "difference" and "ratio". We illustrate these below. Internal habit formation captures the notion that utility is a decreasing function of *ones own* past consumption and marginal utility is an increasing function of own past consumption. Models with external habit emphasize that the operative benchmark is not ones own past consumption but the consumption relative to other agents in the economy.

Constantinides (1990) considers a model with internal habit where utility is defined over the *difference* between current consumption and lagged past consumption. Although the Constantinides (1990) model is in continuous time with a general lag structure, we can illustrate the intuition behind this class of models incorporating "habit" by considering preferences with a one period lag

$$U(c) = E_t \sum_{s=0}^{\infty} \beta^s \frac{(c_{t+s} - \lambda c_{t+s-1})^{1-\alpha}}{1-\alpha} \quad ,\lambda > 0 \tag{30}$$

If λ=1 and the subsistence level is fixed the period utility function specializes to the form $u(c) = \frac{(c-x)^{1-\alpha}}{1-\alpha}$ where x is the fixed subsistence level[32]. The implied local coefficient of relative risk aversion is

$$-\frac{c\,u''}{u©} = \frac{\alpha}{1 - x/c} \tag{31}$$

If x/c =0.8 then the effective risk aversion is 5α!

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. Thus, while this approach cannot resolve the equity premium puzzle without invoking extreme aversion to consumption risk, it can address the risk free rate puzzle since the induced aversion to consumption risk increases the demand for bonds, thereby reducing the risk-free rate. Furthermore, if the growth rate of consumption is assumed to be i.i.d., an implication of

[32] See also the discussion in Weil (1989)

this model is that the risk free rate will vary considerably (and counterfactually) over time. Constantinides (1990) gets around this problem since the growth rate in his model is not i.i.d.[33]

An alternate approach to circumvent this problem has been expounded by Campbell and Cochrane (1999). The model incorporates the possibility of recession as a state variable so that risk aversion varies in a highly nonlinear manner.[34] The risk aversion of investors rises dramatically when the chances of a recession become larger and 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. However, it is an open question whether investors actually have the huge time varying counter-cyclical variations in risk aversion postulated in the model.

Another modification of the Constantinides (1990) approach is to define utility of consumption relative to average per capita consumption. This is an external habit model where preferences are defined over the ratio of consumption to lagged[35] aggregate consumption. Able (1990) terms his model “Catching 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 and this can result in personal consumption falling relative to others. Equity thus becomes an undesirable asset relative to bonds. Since average per capita consumption is rising over time, the

---

[33] In fact, a number of studies suggest that there is a small serial correlation in the growth rate.

[34] If we linearize the “surplus consumption ratio” in the Campbell Cochrane (1999) model, we get the same variation in the risk free rate as in the standard habit model. The nonlinear “surplus consumption ratio” is essential to reducing this variation.

[35] Hence “Catching up with the Joneses” rather than “keeping up with the Joneses.” Abel (1990) footnote 1.

induced demand for bonds with this modification helps in mitigating the risk-free rate puzzle.

Abel (1990) defines utility as the *ratio* of consumption relative to average per capita consumption rather than the difference between the two. It can be shown that this is not a trivial modification.[36] While "difference" habit models can in principle generate a high equity premium, ratio models generate a premium that is similar to that obtained with standard preferences.

To illustrate consider the framework in Abel (1990) specialized to the "catching up with the Joneses" case. At time t, the representative agent in the economy chooses the level of consumption c_t to maximize

$$U(c) = E_t \sum_{t=0}^{\infty} \beta^t \frac{(c_t / C^{\gamma}{}_{t-1})^{1-\alpha}}{1-\alpha}, \alpha > 0 \qquad (32)$$

where C_{t-1} is the lagged aggregate consumption. In equilibrium of course $C_t = c_t$, a fact we use in writing the counter parts of equations (3) and (5) below.

$$1 = \beta E_t \left\{ R_{e,t+1} x_t^{\gamma(\alpha-1)} x_{t+1}^{-\alpha} \right\} \qquad (33)$$

$$1 = \beta R_{f,t+1} E_t \left\{ x_t^{\gamma(\alpha-1)} x_{t+1}^{-\alpha} \right\} \qquad (34)$$

where $x_{t+1} \equiv \frac{c_{t+1}}{c_t}$ is the growth rate of consumption. Under the assumptions made in section

2.1 we can write $R_{f,t+1} = E_t \left\{ x_{t+1}^{\gamma(\alpha-1)} \right\} / \beta E_t \{ x_{t+1}^{-\alpha} \}$ (35)

and $$E_t\{R_{e,t+1}\} = E_t \left\{ x_{t+1}^{\gamma(\alpha-1)} \right\} [E_t\{x_{t+1}\} + A E_t \left\{ x_{t+1}^{1+\gamma(\alpha-1)} \right\}] / A \qquad (36)$$

[36] See Campbell (2001) for a detailed discussion.

We see that in the expression $\ln R_f = -\ln\beta + \alpha\mu_x - 1/2\alpha^2\sigma_x^2 - \gamma(1-\alpha)\mu_x$, the equity premium is $\ln E\{R_e\} - \ln R_f = \alpha\sigma_{x,z}$, which is exactly the same as what was obtained earlier. Hence the equity premium is unchanged! However when $\gamma > 0$, a high α does not lead to the risk free rate puzzle.

The statement, "External habit simply adds a term to the Euler equation 60 which is known at time t, and this does not affect the premium" in Campbell (2001) appears to be inconsistent with the results in Table 1 Panel B in Abel (1990).

Resolution.

Although the "set up" in Abel (1990) and Campbell (2001) is similar, Campbell's result is based on the assumption that asset returns and the growth rate of consumption are jointly log-normally distributed in *both* the "standard time additive" case *and* the "Joneses" case. In Abel (1990) the return distributions are endogenously determined and Campbell's assumption is *internally inconsistent* in the context of that model.

In Abel (1990), with "standard time additive" preferences, *if* consumption growth is log-normally distributed gross asset returns will also be lognormal, however, this is *not* the case with the "Joneses" preferences. In the latter case since $1 + R_{i,t+1} = x_t^{1-\alpha}(x_{t+1} + Ax_{t+1}^{\alpha})/A$, log-normality of x will *not* induce log-normality in $1 + R_{i,t+1}$. ($x_t = c_{t+1}/c_t$).

Abel (1990) reports expressions for $E(1 + R_{i,t+1})$ and $E(1 + R_{f,t+1})$ in equations 17 and 18. Let $\prod_{Abel} = \ln(E(1 + R_{i,t+1})) - \ln(E(1 + R_{f,t+1}))$. In the Abel model with $\theta = 0$ (the "standard time additive" case), *if* the growth rate of consumption is assumed to be lognormally distributed $\prod_{Abel}$ can be written as:

$$\prod_{Abel} = E(\ln(1+R_{i,t+1})) + 0.5Var(\ln(1+R_{i,t+1})) - E(\ln(1+R_{f,t+1})) - 0.5Var(\ln(1+R_{f,t+1})) \qquad (37)$$

or $\prod_{Abel} = \prod_{Campbell} + 0.5[Var(\ln(1+R_{i,t+1})) - Var(\ln(1+R_{f,t+1}))]$ (38)

or $\prod_{Abel} = \prod_{Campbell} + 0.5Var(\ln(x))$. (39)

where $\prod_{Campbell} = E(\ln(1+R_{i,t+1})) - E(\ln(1+R_{f,t+1}))$, is the definition of the equity premium in Campbell (2001).

With "standard time additive" preferences and log-normally distributed returns, the analysis in Abel and Campbell are equivalent. Indeed, a direct evaluation of $\prod_{Abel}$ from equations 17 and 18 in Abel (1990) yields $\prod_{Abel} = \alpha Cov(\ln x, \ln(1+R_i))$. This is identical to that obtained by adjusting equation 62 in Campbell by adding $0.5Var(\ln(x))$.

However, in Abel (1990) with "Joneses" preferences, if the growth rate of consumption is log-normally distributed, asset returns will *not be* lognormal, hence the analysis in Campbell (2001) after equation 60 will *not* apply.

In Abel (1990), as preferences change, return distributions will change, hence if the counterpart of equation 16 (in Campbell) represents the equity premium in the "standard time additive" framework then equation 62 will *not* be the relevant expression for the premium in the "Joneses" case. Counterparts of equations 16 and 62 in Campbell (2001) will not *both* hold *simultaneously* in Abel (1990).

To summarize, models with habit formation and relative or subsistence consumption have had success in addressing the risk free rate puzzle but only limited success with resolving the equity premium puzzle, since in these models effective risk aversion and prudence become implausibly large.

2.3 Idiosyncratic and Uninsurable Income Risk

At a theoretical level, aggregate consumption is a meaningful economic construct if the market is complete, or effectively so.[37] Market completeness is implicitly incorporated into asset pricing models in finance and neoclassical macroeconomics through the assumption of the existence of a representative household. In complete markets, heterogeneous households are able to equalize, state by state, their marginal rate of substitution. The equilibrium in a heterogeneous full-information economy is isomorphic in its pricing implications to the equilibrium in a representative-household, full-information economy, if households have von Neumann-Morgenstern preferences.

Bewley (1982), Mankiw (1986), and Mehra and Prescott (1985) suggest the potential of enriching the asset-pricing implications of the representative-household paradigm, by relaxing the assumption of complete markets.[38]

Current financial paradigms postulate that idiosyncratic income shocks must exhibit three properties in order to explain the returns on financial assets: uninsurability, persistence heteroscedasticity and counter cyclical conditional variance. In infinite horizon models, agents faced with uninsurable income shocks will dynamically self -insure, effectively smoothing consumption. Hence the difference in the equity premium in incomplete markets and complete markets is small.[39]

---

[37] This section draws on Constantinides (2002).

[38] There is an extensive literature on the hypothesis of complete consumption insurance. See, Altonji, Hayashi and Kotlikoff (1992), Attanasio and Davis (1997), Cochrane (1991), and Mace (1991).

[39] Lucas (1994) and Telmer (1993) calibrate economies in which consumers face uninsurable income risk and borrowing or short-selling constraints. They conclude that consumers come close to the complete-markets rule of complete risk sharing, although consumers are allowed to trade in just one security in a frictionless market. Aiyagari and Gertler (1991) and Heaton and Lucas (1996) add transaction costs and/or borrowing costs and reach a similar negative conclusion, provided that the supply of bonds is not restricted to an unrealistically low level.

Constantinides and Duffie (1996), propose a model incorporating heterogeneity that captures the notion that consumers are subject to idiosyncratic income shocks that cannot be insured away. For instance, consumers face the risk of job loss, or other major personal disasters that cannot be hedged away or insured against.[40] Equities and related pro-cyclical investments exhibit the undesirable feature that they drop in value when the probability of job loss increases, as, for instance, in recessions. In economic downturns, consumers thus need an extra incentive to hold equities and other similar investment instruments; the equity premium can then be rationalized as the added inducement needed to make equities palatable to investors.

The model provides an explanation of the counter-cyclical behavior of the equity risk premium: the risk premium is highest in a recession because the stock is the poorest hedge to job loss in a recession. It also provides an explanation of the unconditional equity premium puzzle: even though *per capita* consumption growth is poorly correlated with stocks returns, investors require a hefty premium to hold stocks over short-term bonds because stocks perform poorly in recessions, when an investor is more likely to be laid off.

Since the proposition demonstrates the existence of equilibrium in frictionless markets, it implies that the Euler equations of household (but not necessarily of *per capita*) consumption must hold. Furthermore, since the given price processes have embedded in them whatever predictability of returns of the dividend-price ratios and other instruments that the researcher cares to ascribe to them, the *equilibrium* price processes have this predictability built into them *by construction*.

[40] Storesletten, Telmer, and Yaron (2001) provide empirical evidence from the Panel Study on Income Dynamics (PSID) that idiosyncratic income shocks are persistent and have counter cyclical conditional variance. Storesletten, Telmer, and Yaron (2000) corroborate this evidence by studying household consumption over the life cycle.

Constantinides and Duffie (1996), point out that periods with frequent and large uninsurable idiosyncratic income shocks are associated with both dispersed cross-sectional distribution of the household consumption growth and low stock returns. Brav, Constantinides, and Geczy (2002) provide empirical evidence of the impact of uninsurable idiosyncratic income risk on pricing. They estimate the relative risk aversion (RRA) coefficient and test the set of Euler equations of *household* consumption on the premium of the value-weighted and the equally weighted market portfolio return over the risk-free rate, and on the premium of value stocks over growth stocks. [41]They do not reject the Euler equations of *household* consumption with an economically plausible RRA coefficient of between two and four, although they reject the Euler equations of *per capita* consumption with any value of the RRA coefficient.

Krebs (2000) extends the Constantinides and Duffie (1996) model by introducing rare idiosyncratic income shocks that drive consumption close to zero. In his model, the conditional variance and skewness of the idiosyncratic income shocks are nearly constant over time. He provides a theoretical justification of the difficulty of empirically assessing the contribution of these catastrophic shocks in the low-order cross-sectional moments.

2.4 Models Incorporating a Disaster State and Survivorship Bias

Rietz (1988) has proposed a solution to the puzzle that incorporates a very small probability of a very large drop in consumption. He finds that in such a scenario the risk-free rate is much lower than the return on an equity security. This model requires a 1-in-100

[41] In related studies, Jacobs (1999) studies the PSID database on food consumption; Cogley (1999) and Vissing-Jorgensen (2002) study the CEX database on broad measures of consumption; Jacobs and Wang (2001) study the CEX database by constructing synthetic cohorts; and Ait-Sahalia, Parker, and Yogo (2001) measure the household consumption with the purchases of certain luxury goods.

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 last years for which we have economic 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 very high just after World War II and subsequently declined over time. If real interest rates rose significantly as the war years receded, that evidence would support the Rietz hypothesis. Similarly, if the low probability event precipitating the large decline in consumption were a nuclear war, the perceived probability of such an event has surely varied over the last 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 had moved as predicted, that would support Rietz's disaster scenario. But they did not.

Another attempt at resolving the puzzle proposed by Brown et al (1995) focuses on survival bias.

The central thesis here 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 were unsuccessful and hence the ex-ante equity premium was low. Since it was not known a priori which exchanges would survive, for this explanation to work, stock and bond markets must be differentially impacted by a financial crisis. 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 German hyperinflation, holders of bonds denominated in Reich marks lost virtually all value invested in those assets. During the Poincare′ administration in France during the 1920's, bond-holders lost nearly 90% of the value invested in nominal debt. And in the 1980s, Mexican holders of dollar-denominated debt lost a sizable fraction of its value when the Mexican government, in a period of rapid inflation, converted the debt to pesos and limited the rate at which these funds could be withdrawn. Czarist bonds in Russia and Chinese debt holdings (subsequent to the fall of the Nationalists) suffered a similar fate under communist regimes.

The above examples demonstrate that in times of financial crisis, bonds are as likely to lose value as stocks. Although a survival bias may impact on the *levels* of both the return on equity and debt, there is no evidence to support the assertion that these crises impact *differentially* on the returns to stocks and bonds; hence the equity premium is not impacted. In every instance where trading equity has been suspended, due to political upheavals, etc., governments have either reneged on their debt obligations or expropriated much of the real value of nominal debt through the mechanism of unanticipated inflation.

The difficulty that collectively several model classes have had in explaining the equity premium as a compensation for bearing risk leads us to conclude that perhaps it is *not* a "risk premium" but rather due to other factors. We consider these in the next section.

3. Is The Equity Premium Due to Borrowing Constraints, a Liquidity Premium or Taxes?

3.1 Borrowing Constraints

In models with borrowing constraints and transaction costs, the effect is to force investors to hold an inventory of bonds (precautionary demand) to smooth consumption. Hence in infinite horizon with borrowing constraints, agents come close to equalizing their marginal rates of substitution with little effect on the equity premium.[42] Some recent attempts to resolve the puzzle incorporating both borrowing constraints *and* consumer heterogeneity appear promising. One approach, which departs from the representative agent model, has been proposed in Constantinides, Donaldson and Mehra (2002).

In order to systematically illustrate these ideas, the authors construct an overlapping-generations (OLG) exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period.

The authors explore the implications of a borrowing constraint by deriving and contrasting the stationary equilibria in two versions of the economy. In the *borrowing-constrained* version, the young are prohibited from borrowing and from selling equity short.

[42] This is true unless the supply of bonds is unrealistically low. See Aiyagari and Gertler (1991).

The *borrowing-unconstrained* economy differs from the borrowing-constrained one only in that the borrowing constraint and the short-sale constraint are absent.

An *unconstrained* representative agent's maximization problem is formulated as follows. An agent born in period t solves:

$$\max_{\left\{z^e_{t,i},\, z^b_{t,i}\right\}} E\left(\sum_{i=0}^{2} \beta^i U(C_{t,i}) \right) \tag{40}$$

s.t.

$$c_{t,0} + q^e_t z^e_{t,1} + q^b_t z^b_{t,1} \le w^0 \tag{41}$$

$$c_{t,1} + q^e_{t+1} z^e_{t,2} + q^b_{t+1} z^b_{t,2} \le \left(q^e_{t+1} + d_{t+1} \right) z^e_{t,1} + \left(q^b_{t+1} + b \right) z^b_{t,1} + w^1_{t+1}$$

$$c_{t,2} \le \left(q^e_{t+2} + d_{t+2} \right) z^e_{t,2} + \left(q^b_{t+2} + b \right) z^b_{t,2} \tag{42}$$

$c_{t,j}$ is the consumption in period t +j (j = 0, 1, 2) of a consumer born in period t. There are two types of securities in the model, *bonds,* and *equity* with *ex-coupon* and *ex-dividend* prices q^b_t and q^e_t respectively. Bonds are a claim to a coupon payment b every period, while the equity is a claim to the dividend stream $\{d_t\}$. The consumer born in period t receives deterministic wage income $w^0 > 0$ in period t, when young; stochastic wage income $w^1_{t+1} > 0$ in period t+1, when middle-aged; and zero wage income in period t+2, when old. The consumer purchases $z^e_{t,0}$ shares of stock and $z^b_{t,0}$ bonds when young. The consumer adjusts these holdings to $z^e_{t,1}$ and $z^b_{t,1}$, respectively, when middle-aged. The consumer liquidates his/her entire portfolio when old. Thus $z^e_{t,2} = 0$ and $z^b_{t,2} = 0$.

When considering the borrowing constrained equilibrium the following additional constraints are imposed $z^e_{t,j} > 0$ and $z^b_{t,2} > 0$.

The model introduces two forms of market incompleteness. First, consumers of one generation are prohibited from trading claims against their future wage income with consumers of another generation.[43] Second, consumers of one generation are prohibited from trading bonds and equity with consumers of an unborn generation. As discussed earlier in Section 2.3, absent a complete set of contingent claims, consumer heterogeneity in the form of *uninsurable*, *persistent* and *heteroscedastic* idiosyncratic income shocks, with *counter-cyclical* conditional variance, can potentially resolve empirical difficulties encountered by representative-consumer models.[44]

The novelty of their paper lies in incorporating a life-cycle feature to study asset pricing. The idea is appealingly simple. As discussed earlier, 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 is in states where marginal utility is low. Since 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.[45]

A key insight of their paper is that 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 at his life has uncertain future wage *and* equity income; furthermore, the correlation of equity income with consumption will not be particularly high, as long as stock and wage income are not highly correlated. This is empirically the case, as

---

[43] Being homogeneous within their generation, consumers have no incentive to trade claims with consumers of their own generation.
[44] See Mankiw (1986) and Constantinides and Duffie (1996).

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 the middle-aged. 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, it 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 where the marginal investor finds it unattractive to hold equity. The *deus ex machina* is the *stage* in the life cycle of the marginal investor.

The authors argue that the young, who should be holding equity in an economy without frictions and with complete contraction, are effectively shut out of this market because of borrowing constraints. The young are characterized by low wages; 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 return equity). However, they are prevented from doing so because human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection.

[45] This is precisely the reason as explained earlier why 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

In the presence of borrowing constraints, equity is thus exclusively priced by the middle-aged investors, since the young are effectively excluded from the equity markets and we observe a high equity premium. If the borrowing constraint is relaxed, the young will borrow to purchase equity, thereby raising the bond yield. The increase in the bond yield induces the middle-aged to shift their portfolio holdings from equity to bonds. The increase in demand for equity by the young and the decrease in the demand for equity by the middle-aged work in opposite directions. On balance, the effect 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.

3.2 Liquidity Premium

Bansal and Coleman (1996) develop a monetary model that offers an explanation of the equity premium. In their model, some assets other than money play a key feature by facilitating transactions. This affects the rate of return they offer in equilibrium.

Considering the role of a variety of assets in facilitating transactions, they argue 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 estimate this to be 6 % based on the rate offered on NOW accounts for the period they analyze. Since this is a substantial number, they suggest that other money-like assets may also implicitly include a transaction service component to their return. Insofar as T-bills and equity have a different service component built into their

---

high, i.e. when marginal utility is low, hence their price is (relatively) low and their rate of return high.

returns, this may offer an explanation for the observed equity premium. In fact, if this service component differential were about 5 % there would be no equity premium puzzle.

We argue that this approach can be challenged on three accounts. First, the majority of T-bills are held by institutions, that do not use them as compensatory balances for checking accounts and it is difficult to imagine their having a significant transaction service component. Second, the returns on NOW and other interest bearing accounts have varied over time. These returns have been higher post-1980 than in earlier periods. In fact, for most of the twentieth century, checking accounts were *not* interest bearing. However, contrary to the implications of this model, the equity premium has not diminished in the post-1980 period when presumably the implied transaction service component was the greatest. Third, this model implies that there be a significant yield differential between T-Bills and long term government bonds that presumably do not have a significant transaction service component. However, this has not been the case.

3.3 Taxes and Regulation

McGrattan and Prescott (2000, 2001) take the position that factors other than a premium for bearing non-diversifiable risk account for the large difference in the return on corporate equity and the after-tax real interest rate in the 1960-2000 period. They find that changes in the U.S. tax and legal-regulatory systems that permitted retirement accounts and pension funds to hold corporate equity and reductions in marginal income tax rates account for the high return on corporate equity in this period.

Subsequent to the writing of our equity premium paper (Mehra and Prescott (1985)), real business cycle theory was developed (Kydland and Prescott (1982) and Hansen (1985)). Real business cycle theory uses the stochastic growth model augmented to include the labor-leisure decision. One finding of the real business cycle literature is that the real after-tax interest rate varies in the range from 4 to 4.5 percent. Another finding is that the predicted after-tax return on corporate equity is essentially equal to this real interest rate. These results are closely related to and consistent with what Mehra and Prescott (1985) found in their "Equity Premium Puzzle" paper. The key difference is the empirical counterpart of the real interest rate. Mehra and Prescott (1985) use the highly liquid T-bill rate, corrected for expected inflation. Business cycle theorists (see McGrattan and Prescott (2000, 2001), who build in the details of the tax system) use the intertemporal marginal rate of substitution for consumption to determine this interest rate.

Why was the average real return on T-bills significantly below the real interest rate as found in the business cycle literature? Why was the average real return on corporate equity significantly above this real interest rate in the 1960-2000 period? The low realized real return on T-bills in this period probably has to do with liquidity services T-bills provide. The total T-bill real return, including liquidity services, could very well have been in the range from 4 to 4.5 percent.

A more interesting question is, why was the return on corporate equity so high in the 1960-2000 period? McGrattan and Prescott (2000) answer this question in the process of estimating the fundamental value of the stock market in 1962 and 2000. They chose these two points in time because, relative to GDP, after-tax corporate earnings, net corporate debt, and corporate tangible capital stock were approximately the same and the tax system had been stable for a number of years. Further, at neither point in time was there any fear of full or partial expropriation of capital.

What differed was that the value of the stock market relative to GDP in 2000 was nearly twice as large as in 1962.

What changed between 1962 and 2000 were the tax and legal-regulatory systems. The marginal tax rate on corporate distributions was 43 percent in the 1955-1962 period and only 17 percent in the 1987-2000. This marginal tax rate on dividends does not have consequences for steady-state after-tax earnings or steady-state corporate capital, provided that tax revenues are returned lump-sum to households. This tax rate *does* however have consequences for the value of corporate equity.

The important changes in the legal-regulatory system, most of which occurred in the late 1970s and early 1980s, were that corporate equity was permitted to be held as pension fund reserves and that people could invest on a before tax basis in individual retirement accounts that could include equity. The threat of a lawsuit is why debt assets and not equity with higher returns were held as pension fund reserves in 1962. At that time, little equity was held in defined contribution plans retirement accounts because the total assets in these accounts were then a small number. Thus, debt and not equity could and was held tax free in 1962. In 2000, both can be held tax free in defined benefit and defined contribution pension funds and in individual retirement accounts. Not surprisingly, the assets held in untaxed retirement accounts are large in 2000, being approximately 1.3 GDP (McGrattan and Prescott (2000)).

McGrattan and Prescott (2000, 2001) in determining whether the stock market was overvalued or undervalued vis-a-vis standard growth theory exploit the fact that the value of a set of real assets is the sum of the values of the individual assets in the set. They develop a method for estimating the value of intangible corporate capital, something that is not reported on balance sheets and, like tangible capital, adds to the value of corporations. Their method uses only national account data and the equilibrium condition that after-tax returns are equated across assets. They also incorporate the most important features of the U.S. tax system into the model they use to determine the value of corporate equity, in particular, the fact that capital gains are only taxed upon realization.

The formula they develop for the fundamental value of corporate equities V is

(1) $$V = (1-\tau_d)K_T^{'} + (1-\tau_d)\,(1-\tau_c)\,K_I^{'}$$

where

τ_d is the tax rate on distributions,

τ_c is the tax rate on corporate income,

$K_T^{'}$ is the end-of-period tangible corporate capital stock, and

$K_I^{'}$ is the end-of-period intangible corporate capital stock.

The reasons for the tax factors are as follows. Corporate earnings significantly exceed corporate investment, and as a result, aggregate corporate distributions are large and positive. Historically these distributions have been in the form of dividends. Therefore, the cost of a unit of tangible capital on margin is only $1-\tau_d$ units of forgone consumption. In the case of intangible capital, the consumption cost of a unit of capital is even smaller because investments in intangible capital reduce corporate tax liabilities[46].

The tricky part of the calculation is in constructing a measure of intangible capital. These investments reduce current accounting profits and they increase future economic profits. The formula for steady-state before tax accounting profits is

[46] In fact, formula (1) must be adjusted if economic depreciation and accounting depreciation are not equal and if there is an investment tax credit. See McGrattan and Prescott (2001).

$$(2) \qquad \pi = \frac{i}{1-\tau_c} K_T + i K_I - g K_I ,$$

where g is the steady-state growth rate of the economy and the interest rate i the steady-state after-tax real interest rate. Note that $g K_I$ is steady-state net investment in intangible capital, which reduces accounting profits because it is expensed. Note also, all the variables in formula (2) are reported in the system of national accounts with the exception of i and K_I.

McGrattan and Prescott (2001) estimate i using national income data. Their estimate of i is the after-tax real return on capital in the noncorporate sector, which has as much capital as does the corporate sector. They find that the stock market was neither overvalued nor undervalued in 1962 and 2000. The primary reason for the low valuation in 1962 relative to GDP and high valuation in 2000 relative to GDP is that τ_d was much higher in 1962 than it was in 2000. The secondary reason is that the value of foreign subsidiaries of U.S. corporations grew in the period. An increase in the size of the corporate intangible capital stock was also a contributing factor.

McGrattan and Prescott (2001) find that in the economically and politically stable 1960-2000 period, the after-tax real return on holding corporate equity was as predicted by theory if the changes in the tax and regulatory system were *not* anticipated. These unanticipated changes led to a large unanticipated capital gain on holding corporate equity. Evidence of the importance of these changes is that the share of corporate equity held in retirement accounts and as pension fund reserves increased from essentially zero in 1962 to slightly over 50 percent in 2000. This is important because it means that half of corporate dividends are now subject to zero taxation.

In periods of economic uncertainty, such as those that prevailed in the 1930-1955 period with the Great Depression, World War II, and the fear of another great depression, the survival of the capitalistic system was in doubt. In such times, low equity prices and high real returns on holding equity are not surprising. This is the Brown, Goetzmann, and Ross (1995) explanation of the equity premium. By 1960, the fears of another great depression and of an abandonment of the capitalistic system in the United States had vanished, and clearly other factors gave rise to the high return on equity in the 1960-2000 period.

4. An Equity Premium in the Future?

There is 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 *ex-post* or realized equity premium. This 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 is what we addressed in Mehra and Prescott (1985). However, there is a related concept–the *ex-ante* equity premium. This 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. 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.

Dimson, Marsh and Staunton (2000), Siegel (1998), and Fama and French (2002) document that equity returns over the past 50 years have been higher than their expected values. Fama and French argue that since the average realized return over this period exceeds the one-year ahead conditional forecast (based on the price dividend ratio) by an average of 3.11 to 4.88 percent per year, the expected equity premium should have declined by this amount. The key implication here is that the expected equity premium is small.

If investors have overestimated the equity premium over the second half of this century, Constantinides (2002) argues that "we now have a bigger puzzle on our hands." Why have investors systematically biased their estimates over such a long horizon? He finds no statistical support for the Fama and French claim however.[47]

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 we documented in our 1985 paper 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 as shown earlier (see Figures 1, 2 and 3) it has varied considerably and counter-cyclically over time.

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 either that the historical premium was too high or that the equity premium has diminished.

The data used to document the equity premium over the past 100 years is as good an eco-

---

[47] "Notwithstanding the possibility that regime shifts may well have occurred during this period and that behavior deviations from rationality may have been at work, the simple present-value model matches the gross features of the equity return and the

nomic data set as we have and this is a 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 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 T- bills for investors with a long planning horizon.

price-dividend ratio without having to resort to regime shifts or deviations from rationality." (Constantinides (2002))

References

Abel, Andrew B. “Asset Prices under Habit Formation and Catching Up with the Joneses.” *A.E.R. Papers and Proc.* 80 (May 1990): 38–42

Abel, Andrew B. “Stock Prices Under Time Varying Dividend Risk: An Exact Solution in an Infinite Horizon General Equilibrium Model.” *J. Monetary Econ.* 22 (1988): 375–394.

Abel, Andrew B., N. Gregory Mankiw, Lawrence H. Summers and Richard J. Zeckhauser. “Assessing Dynamic Efficiency: Theory and Evidence.” *Review of Economic Studies* 56 (1989): 1–20.

Ait-Sahalia, Yacine, Jonathan A. Parker, and Motohiro Yogo. “Luxury Goods and the Equity Premium.” Working paper #8417, NBER. 2001

Aiyagari, S. Rao, and Mark Gertler. “Asset Returns with Transactions Costs and Uninsured Individual Risk.” *J. Monetary Econ.* 27 (1991): 311–331.

Altonji, Joseph G., Fumio Hayashi, and Laurence J. Kotlikoff. “Is the Extended Family Altruistically Linked?” *American Economic Review* 82(1992): 1177–1198.

Alvarez, Fernando, and Urban Jermann. “Asset Pricing when Risk Sharing is Limited by Default.” *Econometrica* 48(2000): 775–797.

Attanasio, Orazio P., James Banks, and Sarah Tanner. 2002. “Asset Holding and Consumption Volatility.” *Journal of Political Economy,* forthcoming.

Attanasio, Orazio P., and Steven J. Davis. “Relative Wage Movements and the Distribution of Consumption.” *Journal of Political Economy* 104(1997): 1227–1262.

Auerbach, Allan J., and Laurence J. Kotlikoff. *Dynamic Fiscal Policy*. Cambridge University Press 1987.

Barro, Robert J., and Gary S. Becker. “Population Growth and Economic Growth.” Working paper, Harvard University 1988.

Brock, William A. “An Integration of Stochastic Growth Theory and the Theory of Finance, Part 1: The Growth Model.” in J. Green and J. Scheinkman (eds.), *General Equilibrium, Growth & Trade.* Academic Press, New York 1979.

Bansal, Ravi, and John W. Coleman. “A Monetary Explanation of the Equity Premium, Term Premium and Risk Free Rate Puzzles.” *Journal of Political Economy.* 104 (1996): 1135–1171.

Bansal, Ravi, and Amir Yaron. “Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles.” Working paper #8059, NBER 2000.

Barberis, Nicholas C. “Investing for the Long Run when Returns are Predictable.” *Journal of Finance* 55(2000): 225–264.

Barberis, N., M. Huang, and T. Santos. “Prospect Theory and Asset Prices.” *Quarterly Journal of Economics* 116 (2001): 1–53.

Basak, Suleyman, and Domenico Cuoco. “An Equilibrium Model with Restricted Stock Market Participation.” *The Review of Financial Studies* 11(1998): 309–341.

Benartzi, Shlomo, and Richard H. Thaler. “Myopic Loss Aversion and the Equity Premium Puzzle.” *Quarterly Journal of Economics* 110 (1995): 73–92.

Bewley, Truman F. “Thoughts on Tests of the Intertemporal Asset Pricing Model.” Working paper, Northwestern University, 1982.

Billingsley, Patrick. “Probability and Measure.” John Wiley and Sons. 1995.

Blanchard, Olivier, J. “Movements in the Equity Premium.” *Brooking Papers on Economic Activity* 2 (1993): 75–138.

Boldrin, Michel, Lawrence J. Christiano, and J. D. M. Fisher. “Habit Persistence, Asset Returns, and the Business Cycle.” *American Economic Review* 91 (2001): 149–166.

Brav, Alon, George M. Constantinides, and Christopher C. Geczy. 2002. “Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence.” *Journal of Political Economy*, August, forthcoming.

Brav, Alon, George M. Constantinides, and Christopher C. Geczy. 2001. Asset Pricing and the Cross – sectional Distribution of Household Consumption: A Monte Carlo Investigation.” Work in progress, University of Chicago.

Brav, Alon, and Christopher C. Geczy, 1995. “An Empirical Resurrection of the Simple Consumption CAPM with Power Utility.” Working paper, University of Chicago.

Breeden, D. “An Intertemporal Asset Pricing Model with Stochastic consumption and Investment opportunities.” *Journal of Financial Economics* 7 (1979): 265 – 296.

Brown, S., W. Goetzmann, and S. Ross. “Survival.” *Journal of Finance* 50 (1995): 853 – 873.

Campbell, John Y. 2002, Consumption – based Asset Pricing,” in George M. Constantinides, Milton Harris, and Rene Stulz, eds., *Handbook of the Economics of Finance*. North – Holland Elsevier, Amsterdam, forthcoming.

Campbell, John Y. "Asset Pricing at the Millennium." *Journal of Finance* 55 (2001): 1515–1567.

Campbell, John Y. "Asset Prices, Consumption, and the Business Cycle." Chapter 19 in J.B. Taylor and M. Woodford eds. Handbook of Macroeconomics Vol. 1, (1999), North–Holland, Amsterdam, 1231 – 1303.

Campbell, John Y., and John H. Cochrane. "By Force of Habit: A Consumption – Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy.* 107 (1999): 205–251.

Campbell, John Y., and Robert J. Shiller. "Valuation Ratios and the Long – run Stock Market Outlook." *Journal of Portfolio Management* 24 (1988): 11 – 26.

Cochrane, John H. "A Simple Test of Consumption Insurance." *Journal of Political Economy* 99 (1991): 957–976.

Cochrane, John H. "Where is the Market Going? Uncertain Facts and Novel Theories." *Economic Perspectives* 21 (1997): 3–37.

Cochrane, John H. *Asset Pricing*. Princeton, NJ: Princeton University Press, 2001.

Cochrane, John H., and Lars Peter Hansen. "Asset Pricing Explorations for Macroeconomics," in Olivier J. Blanchard and Stanley Fischer, eds., NBER, *Macroeconomics Annual.* Cambridge, Mass.: M.I.T. Press, 1992.

Cogley, Timothy. "Idiosyncratic Risk and the Equity Premium: Evidence from the Consumer Expenditure Survey." Working paper, Arizona State University, 1999.

Constantinides, George M. 2002. "Rational Asset Prices." *Journal of Finance*. Forthcoming.

Constantinides, George M. "Habit Formation: A Resolution of the Equity Premium Puzzle." *Journal of Political Economy* 98 (1990): 519–543.

Constantinides, George M., Donaldson, John B., and Rajnish Mehra. "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle." *Quarterly Journal of Economics* 118 (2002): 269–296.

Constantinides, George M., and Darrell Duffie. "Asset Pricing with Heterogeneous Consumers." *Journal of Political Economy* 104 (1996): 219–240.

Cowles, Alfred, and Associates. "Common Stock Indexes." 2d ed. Cowles Commission Monograph no.3. Bloomington, Ind.: Principia Press, 1939.

Cox, John C., Jonathan E.Ingersoll, Jr., and Stephen A.Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53 (1985):385–407.

Daniel, Kent, and David Marshall. "The Equity Premium Puzzle and the Risk-free Rate Puzzle at Long Horizons." *Macroeconomic Dynamics* l (1997): 452–484.

Danthine, Jean-Pierre, John B. Donaldson, and Rajnish Mehra. "The Equity Premium and the Allocation of Income Risk." *Journal of Economic Dynamics and Control* 16 (1992): 509 – 532.

Danthine, Jean-Pierre, and John B. Donaldson. "Intermediate Financial Theory." Upper Saddle River, NJ: Prentice Hall, 2001.

Debreu,G.," Valuation Equilibrium and Pareto Optimum." *Proceedings of the National Academy of Sciences* 70 (1954) 588-592.

Donaldson, John B., and Rajnish Mehra. "Stochastic Growth with Correlated Production Shocks." *Journal of Economic Theory.* 29 (1983): 282–312.

Donaldson, John B., and Rajnish Mehra. "Comparative Dynamics of an Equilibrium Intertemporal Asset Pricing Model." *Review of Economic Studies* 51(1984): 491–508.

Donaldson, John B. Thore H. Johnsen and Rajnish Mehra. "On the Term Structure of Interest Rates." *Journal of Economic Dynamics and Control* 14 (1990): 571–596.

Davis, Stephen J., and Paul Willen. "Using Financial Assets to Hedge Labor Income Risk: Estimating the Benefits." Working paper, University of Chicago, 2000.

Dimson, E., Paul Marsh, and Michael Staunton. "The Millennium Book: A Century of Investment Returns." Working paper, ABN Amro and London Business School, 2000.

Duffie, Darrell. *Dynamic Asset Pricing Theory.* 3d edition. Princeton, NJ: Princeton University Press. 2001.

Epstein, Larry G., and Stanley E. Zin, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis." *Journal of Political Economy* 99 (1991): 263 – 86.

Fama, Eugene F., and Kenneth R. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics* 22(1988): 3–25.

Fama, Eugene F., and Kenneth R. French. "The Equity Premium." *Journal of Finance* 57(April 2002), forthcoming.

Ferson, Wayne E., and George M. Constantinides. "Habit Persistence and Durability in Aggregate Consumption." *Journal of Financial Economics* 29(1991): 199–240.

Gabaix, Xavier, and David Laibson. "The 6D Bias and the Equity Premium Puzzle," in Ben Bernanke and Ken Rogoff, eds., *NBER Macroeconomics Annual 2001.* Cambridge, Mass: MIT Press.

Grossman, S. J. and R. J. Shiller (1981), "The Determinants of the Variability of Stock Market Prices," American Economic Review 71:222 - 227.

Hansen, Lars Peter, and Ravi Jagannathan. "Implications of Security Market Data for Models of Dynamic Economies." *Journal of Political Economy* 99 (April 1991): 225–62.

Hansen, Lars Peter, and Singleton, Kenneth J. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica* 50 (1982): 1269-88.

He, Hua, and David M. Modest. "Market Frictions and Consumption-based Asset Pricing." *Journal of Political Economy* 103 (1995): 94 – 117.

Heaton, John. "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications." *Econometrica* 66 (1995): 681–717.

Heaton, John, and Deborah J. Lucas. "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing." *Journal of Political Economy* 104 (1996): 443–487.

Heaton, John, and Deborah J. Lucas. "Market Frictions, Savings Behavior and Portfolio Choice." *Journal of Macroeconomic Dynamics* 1 (1997): 76–101.

Heaton, John C., and Deborah J. Lucas. "Stock Prices and Fundamentals," in Ben Bernanke and Julio Rotemberg, eds., *Macroeconomics Annual 1999,* National Bureau of Economic Research. Cambridge, Mass: MIT Press, 1999.

Heaton, John, C., and Deborah J. Lucas. "Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk." *Journal of Finance* 55(2000).

Homer, S., *A History of Interest Rates.* New Brunswick, NJ: Rutgers University Press. 1963.

Ibbotson Associates. "Stocks, Bonds, Bills and Inflation". 2000 Yearbook. Chicago: Ibbotson Associates, 2001.

Jacobs, Kris. "Incomplete Markets and Security Prices: Do Asset-pricing Puzzles Result from Aggregation Problems?" *Journal of Finance* 54(1999): 123–163.

Jacobs, Kris, and Kevin Q. Wang. "Idiosyncratic Consumption Risk and the Cross-section of Asset Returns." Working paper, McGill University and University of Toronto, 2001.

Kandel, S. and R.F. Stambaugh, "Asset Returns and Intertemporal Preferences." *Journal of Monetary Economics* 27 (1991): 39–71.

Kocherlakota, Narayana R. "The Equity Premium: It's Still a Puzzle." *J. Econ. Literature*, 34 (1996): 42–71.

Krebs,Tom. "Consumption-based Asset Pricing with Incomplete Markets." Working paper, Brown University, 2000.

Kydland, Finn, and Edward C. Prescott. "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (1982): 1345–1371.

LeRoy, Stephen H, and Jan Werner. *Principles of Financial Economics.* New York, NY: Cambridge University Press, 2001.

LeRoy, S.F. and R.D. Porter. "The Present-Value Relation: Tests Based upon Implied Variance Bounds." Econometrica 49 (1981): 555-574.

Litterman, R.B. "Bayesian Procedure for Forecasting with Vector Auto-regressions."Working paper, MIT, 1980.

Lucas, Deborah J. "Asset Pricing with Undiversifiable Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle." *Journal of Monetary Economics* 34 (1994): 325–341.

Lucas, Robert E., Jr. "Asset Prices in an Exchange Economy." *Econometrica* 46(1978): 1429–45.

Luttmer, Erzo G. J. " Asset Pricing in Economies with Frictions." *Econometrica* 64(1996): 1439 –1467.

Lynch, Anthony W. "Decision Frequency and Synchronization Across Agents: Implications for Aggregate Consumption and Equity Returns." *Journal of Finance* 51(1996): 1479–1497.

Mace, Barbara J. "Full Insurance in the Presence of Aggregate Uncertainty." *Journal of Political Economy* 99 (1991): 928–956.

Macaulay, F.R. " The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856." New York. National Bureau of Economic Research (1938).

Mankiw, N.G. "The Equity Premium and the Concentration of Aggregate Shocks." *Journal of Financial Economics* 17(1986): 211–219.

Mankiw, N. Gregory, and Stephen P. Zeldes. "The Consumption of Stockholders and Nonstockholders." *Journal of Financial Economics* 29 (1991): 97–112.

McGrattan, Ellen R., and Edward C. Prescott. "Is the Market Overvalued?" *Federal Reserve Bank of Minneapolis Quarterly Review* 24 (2000): 20–40.

McGrattan, Ellen R., and Edward C. Prescott. “Taxes, Regulations, and Asset Prices.” Working paper 610, Federal Reserve Bank of Minneapolis, 2001.

Mehra, Rajnish, “On the Existence and Representation of Equilibrium in an Economy with Growth and Nonstationary Consumption,” *International Economic Review*, 29 (February 1988): 131–35.

Mehra, Rajnish. “On the Volatility of Stock Prices: An Exercise in Quantitative Theory”. *International Journal of Systems Science* 29(1998): 1203–11.

Mehra, Rajnish. “The Equity Premium: Why Is It A Puzzle” *Financial Analysts Journal* January/February (2003) :54-69

Mehra, Rajnish, and Edward C.Prescott. “The Equity Premium: A Puzzle.” *J. Monetary Econ.* 15 (1985): 145–61.

Mehra, Rajnish, and Edward C. Prescott. “The Equity Premium: A Solution?” *J. Monetary Econ.* 22 (1988): 133–36.

Mehra, Rajnish, and Raaj Sah. 2002 “Mood Fluctuations, Projection Bias, and Volatility of Equity Prices.” *Journal of Economic Dynamics and Control.* Forthcoming

Merton, Robert C. “Optimum Consumption and Portfolio Rules in a Continuous Time Model.” *Journal of Theory* 3 (1971): 373 – 413.

Pastor, Lubos, and Robert F. Stambaugh. “The Equity Premium and Structural Breaks”. *Journal of Finance* 56 (2001):1207 – 1239

Prescott, Edward C., and Rajnish Mehra. “Recursive Competitive Equilibrium: The Case of Homogeneous Households.” *Econometrica* 48 (1980): 1365 –1379.

Rietz, Thomas A. “The Equity Risk Premium: A Solution.” *J. Monetary Econ.* 22 (1988): 117 – 131.

Rubinstein, Mark. “The Valuation of Uncertain Income Streams and the Pricing of Options.” *Bell Journal of Economics* 7 (1976): 407 – 425.

Schwert, G. William. “Indexes of U.S. Stock Prices from 1802 to 1987.” *Journal of Business* 63 (1990): 399 – 426.

Smith, W.B., and A.H. Cole, “Fluctuations in American Business, 1790 – 1860.” Cambridge, Mass: Harvard University Press, 1935.

Storesletten, Kjetil, Chris I.Telmer,and Amir, Yaron. “Asset Pricing with Idiosyncratic Risk and Overlapping Generations.” Working paper.: Carnegie Mellon University, Pittsburgh,Pa 1999.

Storesletten, Kjetil, Chris I. Telmer, and Amir, Yaron. “Consumption and Risk Sharing over the Lifecycle”. Working paper, Carnegie Mellon University, Pittsburg,Pa. 2000.

Siegel, Jeremy. *Stocks for the Long Run*. 2d ed. New York: Irwin. 1998.

Telmer, Chris I. “Asset – pricing Puzzles and Incomplete Markets.” *Journal of Finance* 49 (1993): 1803 – 1832.

Vissing-Jorgensen, Annette. 2002. “Limited Asset Market Participation and the Elasticity of Intertemporal Substitution.” *Journal of Political Economy*. forthcoming.

Weil, Philippe. “The Equity Premium Puzzle and the Risk - Free Rate Puzzle.” *J. Monetary Econ*. 24 (November 1989): 401 – 421.

Appendix A

Suppose the distribution of returns period by period is independently and identically distributed. Then as the number of periods tends to infinity, the future value of the investment, computed at the arithmetic average of returns tends to the expected value of the investment with probability 1.

To see this,
let $V_T = \prod_{t=1}^{T}(1+r_t)$, where r_t is the asset return in period t and V_T is the terminal value of one dollar at time T.

Then $E(V_T) = E\left[\prod_{i=1}^{T}(1+r_t)\right]$. Since the r_ts are assumed to be uncorrelated, we have

$$E(V_T) = \prod_{i=1}^{T} E(1+r_t).$$

or $$E(V_T) = \prod_{i=1}^{T}(1+E(r_t)).$$

Let the arithmetic average, $AA = \frac{1}{T}\sum_{t=t}^{T} r_t$

Then by the strong law of large numbers (Theorem 22.1, Billingsley)

$$E(V_T) \to \prod_{i=1}^{T}(1+AA) \text{ as } T \to \infty$$

or $E(V_T) \to (1+AA)^T$ as the number of periods T becomes large.

If asset returns, r_t, are identically and independently log normally distributed, then, as the number of periods tends to infinity, the future value of an investment compounded at the continuously compounded geometric average rate tends to the median value of the investment.

Let $V_T = \prod_{t=1}^{T}(1+r_t)$, where r_t is the asset return in period t and V_T is the terminal value of one dollar at time T.

The Geometric Average is defined by:

$$GA = \left[\prod_{t=1}^{T}(1+r_t)\right]^{1/T} - 1$$

hence $V_T = (1+GA)^T$ and $\ln(1+GA) = \frac{1}{T}\sum \ln(1+r_t)$

Let the continuously compounded geometric rate of return $= \mu_{rc}$. Then by definition

$$\ln(1+GA) = \mu_{rc}$$

or $\quad 1+GA = e^{\mu_{rc}}$

and $\quad (1+GA)^T = e^{T\mu_{rc}}$

By the properties of the lognormal distribution the median value of $V_T = e^{E(\ln V_T)}$ and by the strong law of large numbers $E(\ln V_T) = \sum E\ln(1+r_t) \rightarrow T\mu_{re}$ as $T \rightarrow \infty$ (Theorem 22.1, Billingsley).

Hence the median value of $V_T = e^{T\mu_{rc}} = (1+GA)^T$ as claimed above.

Appendix B

The Original Analysis of the Equity Premium Puzzle

In this Appendix we present our original analysis of the equity premium puzzle. Needless to say, it draws heavily from Mehra and Prescott (1985).

The economy, asset prices and returns

We employ a variation of Lucas' (1978) pure exchange model. Since per capita consumption has grown over time, we assume that the growth rate of the endowment follows a Markov process. This is in contrast to the assumption in Lucas' model that the endowment level follows a Markov process. Our assumption, which requires an extension of competitive equilibrium theory, enables us to capture the non-stationarity in the consumption series associated with the large increase in per capita consumption that occurred in the 1889-1978 period.

The economy we consider was judiciously selected so that the joint process governing the growth rates in aggregate per capita consumption and asset prices would be stationary and easily determined. The economy has a single representative 'stand-in' household. This unit orders its preferences over random consumption paths by

$$E_0\left\{\sum_{t=0}^{\infty} \beta^t U(c_t)\right\}, \qquad 0 < \beta < 1, \tag{1B}$$

where c_t is per capita consumption, β is the subjective time discount factor, $E\{\cdot\}$ is the expectation operator conditional upon information available at time zero (which denotes the present time) and $U: R_+ \rightarrow R$ is the increasing concave utility function. To insure that the equilibrium return process is stationary, we further restrict the utility function to be of the constant relative risk aversion (CRRA) class

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

The parameter α measures the curvature of the utility function. When α is equal to one, the utility function is defined to be the logarithmic function, which is the limit of the above function as α approaches one.

We assume there is one productive unit which produces output y_t in period t which is the period dividend. There is one equity share with price p_t that is competitively traded; it is a claim to the stochastic process $\{y_t\}$.

The growth rate in y_t is subject to a Markov chain; that is,

$$y_{t+1} = x_{t+1} y_t \tag{3B}$$

where $x_{t+1} \in \{\lambda_1,...,\lambda_n\}$ is the growth rate, and $\Pr\{x_{t+1} = \lambda_1; x_t = \lambda_j\} = \phi_{ij}$. (4B)

It is also assumed that the Markov chain is ergodic. The λ_i are all positive and $y_0 > 0$. The random variable y_t is observed at the beginning of the period, at which time dividend payments are made. All securities are traded ex-dividend. We also assume that the matrix A with elements $a_{ij} \equiv \beta\phi_{ij}\lambda_j^{1-\alpha}$ for i, j = 1,…,n is stable; that is, lim A^m as m → ∞ is zero. In Mehra (1988) it is shown that this is necessary and sufficient for expected utility to exist if the stand-in household consumes y_t every period. The paper also defines and establishes the existence of a Debreu (1954) competitive equilibrium with a price system having a dot product representation under this condition.

Next we formulate expressions for the equilibrium time t price of the equity share and the risk-free bill. We follow the convention of pricing securities ex-dividend or ex-interest payments at time t, in terms of the time t consumption good. For any security with process $\{d_s\}$ on payments, its price in period t is

$$P_t = E_t\left\{\sum_{s=t+1}^{\infty} \beta^{s-t} U'(y_s) d_s / U'(y_t)\right\}, \tag{5B}$$

as the equilibrium consumption is the process $\{y_s\}$and the equilibrium price system has a dot product representation.

The dividend payment process for the equity share in this economy is $\{y_s\}$. Consequently, using the fact that $U'(c) = c^{-\alpha}$,

$$\begin{aligned} P_t^e &= P^e(x_t, y_t) \\ &= E\left\{\sum_{s=t+1}^{\infty} \beta^{s-t} \frac{y_t^{\alpha}}{y_s^{\alpha}} y_s \middle| x_t, y_t\right\} \end{aligned} \tag{6B}$$

Variables $x_t$and $y_t$are sufficient relative to the entire history of shocks up to, and including, time t for predicting the subsequent evolution of the economy. They thus constitute legitimate state variables for the model. Since $y_s = y_t x_{t+1} \ldots x_s$, the price of the equity security is homogeneous of degree one in y_t which is the current endowment of the consumption good. As the equilibrium values of the economies being studied are time invariant functions of the state (x_t, y_t), the subscript t can be dropped. This is accomplished by redefining the state to be the pair (c,i), if $y_t = c$ and $x_t = \lambda_i$. With this convention, the price of the equity share from (6B) satisfies

$$p^e(c,i) = \beta \sum_{j=1}^{n} \phi_{ij} (\lambda_j c)^{-\alpha} [p^e(\lambda_j c, j) + \lambda_j c] c^{\alpha}. \tag{7B}$$

Using the result that $p^e(c,i)$ is homogeneous of degree one in c, we represent this function as

$$p^e(c,i) = \mathrm{w}_i c \tag{8B}$$

where w_i is a constant. Making this substitution in (7B) and dividing by c yields

$$w_i = \beta \sum_{j=1}^{n} \phi_{ij} \lambda_j^{(1-\alpha)} (w_j + 1) \qquad \text{for i = 1,…, n.} \tag{9B}$$

This is a system of n linear equations in n unknowns. The assumption that guaranteed existence of equilibrium guarantees the existence of a unique positive solution to this system.

The period return if the current state is (c, i) and next period state (λ_jc, j) is

$$\begin{aligned} r_{ij}^e &= \frac{p^e(\lambda_j c, j) + \lambda_j c - p^e(c,i)}{p^e(c,i)} \\ &= \frac{\lambda_j (w_j + 1)}{w_i} - 1 \ , \end{aligned} \tag{10B}$$

The equity's expected period return if the current state is i is

$$R_i^e = \sum_{j=1}^{n} \phi_{ij} r_{ij}^e \ . \tag{11B}$$

Capital letters are used to denote expected return. With the subscript i, it is the expected return conditional upon the current state being (c, i). Without this subscript it is the expected return with respect to the stationary distribution. The superscript indicates the type of security.

The other security considered is the one-period real bill or riskless asset, which pays one unit of the consumption good next period with certainty.

From (6B),

$$p_i^f = p^f(c,i)$$

$$= \beta \sum_{j=1}^{n} \phi_{ij} U'(\lambda_j c) / U'(c) \tag{12B}$$

$$= \beta \sum \phi_{ij} \lambda_j^{-\alpha}$$

The certain return on this riskless security is

$$R_i^f = 1 / p_i^f - 1 \ , \tag{13B}$$

when the current state is (c, i).

As mentioned earlier, the statistics that are probably most robust to the modeling specification are the means over time. Let $\pi \in R^n$ be the vector of stationary probabilities on i. This exists because the chain on i has been assumed to be ergodic. The vector π is the solution to the system of equations

$$\pi = \phi^T \pi \ ,$$

with

$$\sum_{i=1}^{n} \pi_i = 1 \text{ and } \phi^T = \{\phi_{ji}\}.$$

The expected returns on the equity and the risk-free security are, respectively,

$$R^e = \sum_{i=1}^{n} \pi_i R_i^e \text{ and } R^f = \sum_{i=1}^{n} \pi_i R_i^f \ . \tag{14B}$$

Time sample averages will converge in probability to these values given the ergodicity of the Markov chain. The risk premium for equity is $R^e - R^f$, a parameter that is used in the test.

The parameters defining preferences are α and β while the parameters defining technology are the elements of $[\phi_{ij}]$ and $[\lambda_i]$. Our approach is to assume two states for the Markov chain and to restrict the process as follows:

$$\lambda_1 = 1 + \mu + \delta, \qquad \lambda_2 = 1 + \mu - \delta,$$

$$\phi_{11} = \phi_{22} = \phi, \qquad \phi_{12} = \phi_{21} = (1 - \phi).$$

The parameters μ, ϕ, and δ now define the technology. We require $\delta > 0$ and $0 < \phi < 1$. This particular parameterization was selected because it permitted us to independently vary the average growth rate of output by changing μ, the variability of consumption by altering δ, and the serial correlation of growth rates by adjusting ϕ.

The parameters were selected so that the average growth rate of per capita consumption, the standard deviation of the growth rate of per capita consumption and the first-order serial correlation of this growth rate, all with respect to the model's stationary distribution, matched the sample values for the U.S. economy between 1889-1978. The sample values for the U.S. economy were 0.018, 0.036 and –0.14, respectively. The resulting parameter's values were $\mu = 0.018$, $\delta = 0.036$ and $\phi = 0.43$. Given these values, the nature of the test is to search for parameters α and β for which the model's averaged risk-free rate and equity risk premium match those observed for the U.S. economy over this ninety-year period.

The parameter α, which measures peoples' willingness to substitute consumption between successive yearly time periods is an important one in many fields of economics. As mentioned in the text there is a wealth of evidence from various studies that the coefficient of risk aversion α is a small number, certainly less than 10. A number of these studies are documented in Mehra and Prescott (1985). This is an important restriction, for with large α virtually any pair of average equity and risk-free returns can be obtained by making small changes in the process on consumption.

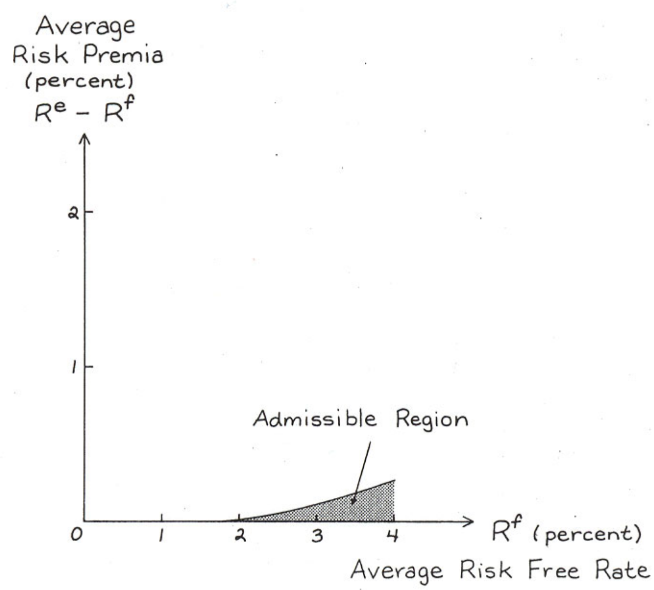


Figure 1B. Set of admissible average equity risk premia and real returns

Given the estimated process on consumption, Figure 1B depicts the set of values of the average risk-free rate and equity risk premium which are both consistent with the model and result in average real risk-free rates between zero and four percent. These are values that can be obtained by varying preference parameters α between zero and ten and β between zero and one.

The observed real return of 0.80 percent and equity premium of 6 percent is clearly inconsistent with the predictions of the model. The largest premium obtainable with the model is 0.35 percent, which is not close to the observed value.

An advantage of our approach is that we can easily test the sensitivity of our results to such distributional assumptions. With α less than ten, we found that our results were essentially unchanged for very different consumption processes, provided that the mean and variances of growth rates equaled the historically observed values. We use this fact in motivating the discussion in the text.