now
the essence of knowledge

The Equity Premium Puzzle: A Review

Rajnish Mehra

Department of Economics, University of California, Santa Barbara and NBER, mehra@econ.ucsb.edu

Abstract

Over two decades ago, Mehra and Prescott (1985) challenged the finance profession with a poser: the historical US equity premium is an order of magnitude greater than can be rationalized in the context of the standard neoclassical paradigm of financial economics. This regularity, dubbed "the equity premium puzzle," has spawned a plethora of research efforts to explain it away. In this review, the author takes a retrospective look at the original paper and explains the conclusion that the equity premium is not a premium for bearing non-diversifiable risk.

1

Introduction

Over two decades ago, Edward Prescott and I (Mehra and Prescott, 1985) challenged the profession with a poser: the historical US equity premium, (the return earned by a risky security in excess of that earned by a relatively risk free US T-bill) is an order of magnitude greater than can be rationalized in the context of the standard neoclassical paradigm of financial economics. This regularity, dubbed "the equity premium puzzle," has spawned a plethora of research efforts to explain it away. In this review, I take a retrospective look at our original paper and show why we concluded that the equity premium is not a premium for bearing non-diversifiable risk.[1] I provide a birds eye view of the vast literature spawned by our paper and touch on other issues that may be of interest to the researcher who did not have a ringside seat over the last 25 years. The reader is referred to Mehra (2007, 2008) and the papers therein for a detailed survey.

The year 1978 saw the publication of Robert Lucas's seminal paper "Asset Prices in an Exchange Economy." Its publication transformed

---

[1] This article draws on material in Mehra (2003), Mehra and Prescott (2003, 2007, 2008a,b), and Donaldson and Mehra (2008). Some sections of this article closely follow the exposition in these papers. The acknowledgments in these papers continue to apply.

asset pricing and substantially raised the level of discussion, providing a theoretical construct to study issues that could not be addressed within the dominant paradigm at the time, the Capital Asset Pricing Model.[2] A crucial input parameter for using the latter is the equity premium (the return earned by a broad market index in excess of that earned by a relatively risk-free security). Lucas' asset pricing model allowed one to pose questions about the magnitude of the equity premium. In our paper "The Equity Premium: A Puzzle"[3] we decided to address this issue.

This review is organized into a further five sections. Section 2 documents the historical equity premium in the United States and in selected countries with significant capital markets (in terms of market value). Section 3 examines the question, "Is the equity premium a premium for bearing non-diversifiable risk?" Section 4 addresses risk and preference based explanations of the equity premium. Section 5, in contrast, reviews the nascent literature that takes as given the findings in Mehra and Prescott (1985) and tries to account for the equity premium by *factors other than aggregate risk*. Section 6 concludes the review.

---

[2] See Mossin (1966) for a lucid articulation.
[3] Mehra and Prescott (1985).

2

Empirical Facts

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 2.1, which reports the unconditional estimates[1] for the US equity premium based on the various data sets used in the literature, going back to 1802. The average annual real return, (the inflation adjusted return) on the US stock market over the last 116 years has been about 7.67 percent. Over the same period, the return on a relatively riskless security was a paltry 1.31 percent. The difference between these two returns, the "equity premium," was 6.36 percent.

Furthermore, this pattern of excess returns to equity holdings is not unique to the United States but is observed in every country with a significant capital market. The United States together with the

---

[1] To obtain unconditional estimates we use the entire data set to form our estimate. The Mehra–Prescott data set spans 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. Constantinides (2002) makes a distinction between expected returns and sample average returns and provides a statistical correction. This turns out to be minor, thereby reinforcing the puzzle.

Table 2.1 US equity premium using different data sets.

Data set	Real return on a market index (%) Mean	Real return on a relatively riskless security (%) Mean	Equity premium (%) Mean
1802–2004 (Siegel)	8.38	3.02	5.36
1871–2005 (Shiller)	8.32	2.68	5.64
1889–2005 (Mehra–Prescott)	7.67	1.31	6.36
1926–2004 (Ibbotson)	9.27	0.64	8.63

United Kingdom, 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 7.4 percent over the last 106 years, an impressive 6.1 percent premium over the average bond return of 1.3 percent. Similar statistical differentials are documented for France, Germany, and Japan. Table 2.2 illustrates the equity premium for these countries.

The dramatic investment implications of this differential rate of return can be seen in Table 2.3, which maps the capital appreciation of $1 invested in different assets from 1802 to 2004 and from 1926 to 2004.

One dollar invested in a diversified stock index yields an ending wealth of $655,348 versus a value of $293, in real terms, for one dollar invested in a portfolio of T-bills for the period 1802–2004. The

Table 2.2 Equity premium for selected countries.

Country	Period	Mean real return		Equity premium (%)
		Market index (%)	Relatively riskless security (%)	
United Kingdom	1900–2005	7.4	1.3	6.1
Japan	1900–2005	9.3	−0.5	9.8
Germany	1900–2005	8.2	−0.9	9.1
France	1900–2005	6.1	−3.2	9.3
Sweden	1900–2005	10.1	2.1	8.0
Australia	1900–2005	9.2	0.7	8.5
India	1991–2004	12.6	1.3	11.3

Sources: Dimson et al. (2002) and Mehra (2007) for India.

Table 2.3 Terminal value of $1 invested in stocks and bonds.

Investment period	Stocks		T-bills	
	Real	Nominal	Real	Nominal
1802–2004	$655,348	$10,350,077	$293	$4,614
1926–2004	$238.30	$2,533.43	$1.54	$17.87

Sources: Ibbotson (2005) and Siegel (2002)

corresponding values for the 78-year period, 1926–2004, are $238.30 and $1.54. It is assumed 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 Trust funds in the stock market.
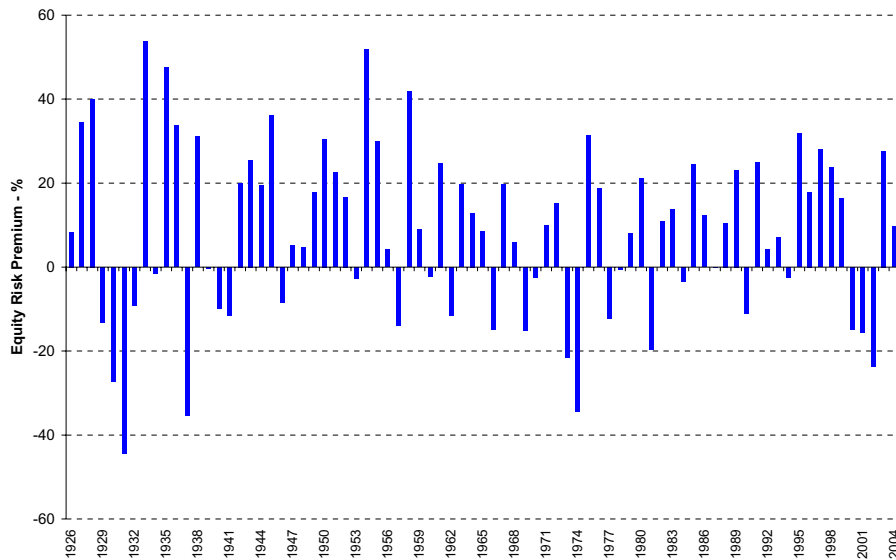

Fig. 2.1 Realized equity risk premium per year: 1926–2004.
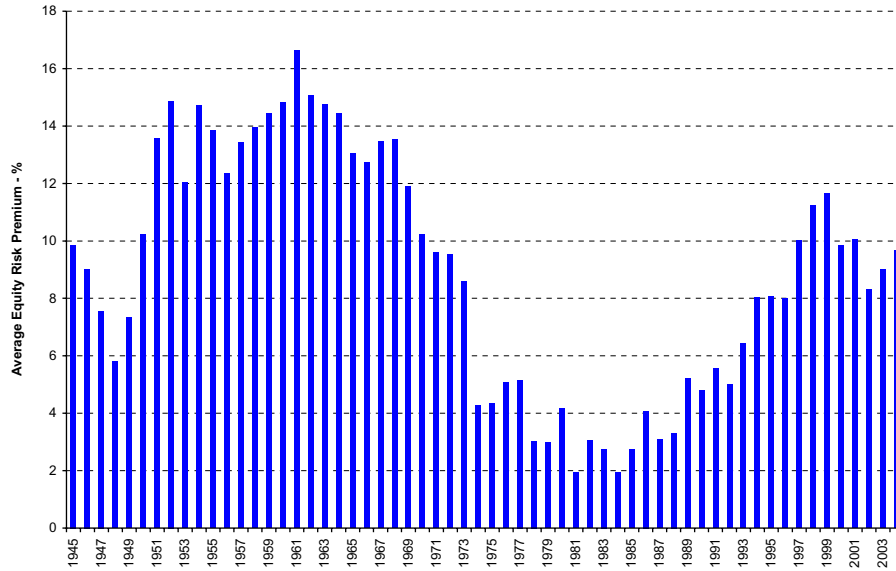*Source*: Ibbotson (2006).

Fig. 2.2 Equity risk premium over 20-year periods: 1926–2004.
Source: Ibbotson (2006).

2.1   Variation in the Equity Premium Over Time

The equity premium has varied considerably over time, as illustrated
in Figures 2.1 and 2.2. Furthermore, the variation depends on the time
horizon over which it is measured. There have even been periods when
it has been negative.

3

Is the Equity Premium a Premium for Bearing Non-diversifiable Risk?

Why have stocks been such an attractive investment relative to bonds? Why has the rate of return on stocks been higher than that 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 3.1 and 3.2 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 1889–2005.[1]

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

---

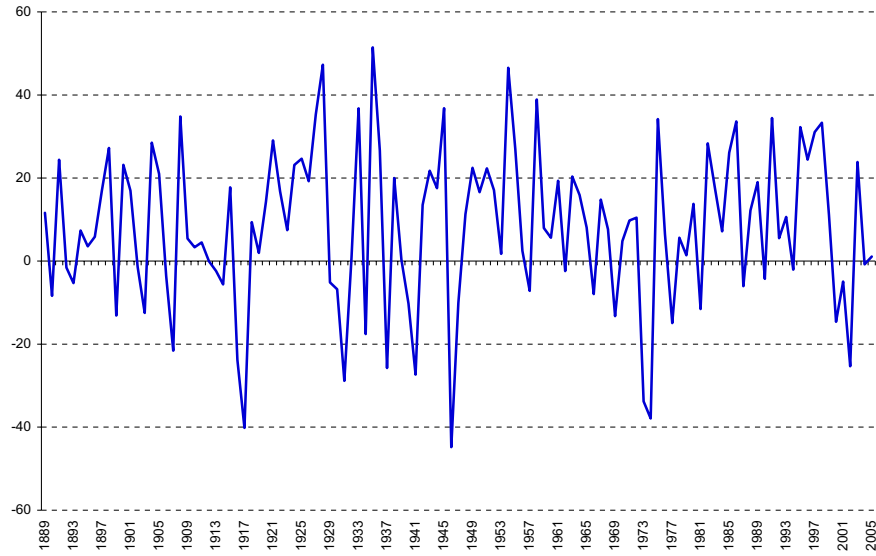[1] The index did not consist of 500 stocks for the entire period.

Fig. 3.1 Real annual return on the S&P 500 index — %: 1889–2005.
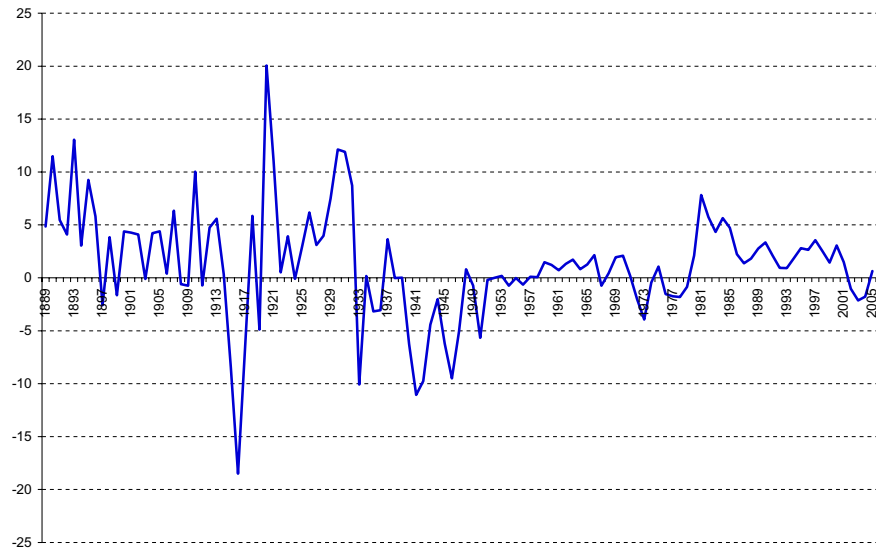*Source*: Mehra and Prescott (1985). Data updated by the authors.

Fig. 3.2 Real annual return on T-bills — %: 1889–2005.
Source: Mehra and Prescott (1985). Data updated by the authors.

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 their low rates of return.

To return to the original question: are stocks that much riskier than T-bills so as to justify a 7 percentage differential in their rates of return?

What came as a surprise to many economists and researchers in finance was the conclusion of our paper, 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 1 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).

3.1 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), Becker and Barro (1988), Brock (1979), Cox et al. (1985), Donaldson and Mehra (1984), Kydland and Prescott (1982), Lucas (1978), 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 peoples' 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 our 1985 paper.[2]

To illustrate this we employ a variation of Lucas' (1978) pure exchange model rather than the production economy studied in Prescott and Mehra (1980). This is an appropriate abstraction if we use the equilibrium relation between the consumption and asset returns to estimate the premium for bearing non-diversifiable risk. Introducing production would add additional restrictions on the process for consumption. In a production economy, consumption would be endogenously determined, restricting the class of consumption processes that could be considered; hence the set of equilibria in a production economy will be a subset of those in an exchange economy. It follows that if the equity premium cannot be accounted for in an exchange economy, modifying the technology to incorporate production will not alter this conclusion.[3] To examine the role of other factors on mean asset returns, it would be necessary to introduce other features of reality such as taxes and intermediation costs as has recently been done.[4] If the model had accounted for differences in average asset returns, the next step would have been to use the neoclassical growth model, which has production, to see if this abstraction accounted for the observed large differences in average asset returns.

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,[5] 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.

---

[2] The reader is referred to McGrattan and Prescott (2003) and Mehra and Prescott (2007, 2008b) for an alternative perspective.

[3] See the Appendix for a proof.

[4] See McGrattan and Prescott (2003) and Mehra and Prescott (2007, 2008a)

[5] This was accomplished in Mehra (1988).

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

$$E_0 \left\{ \sum_{t=0}^{\infty} \beta^t U(c_t) \right\}, \quad 0 < \beta < 1, \tag{3.1}$$

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.[6]

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 expectations operator conditional upon information available at time zero, (which denotes the present time) and U: $R_+ \to R$ is the increasing, continuously differentiable concave utility function. We further restrict the utility function to be of the constant relative risk aversion (CRRA) class

$$U(c, \alpha) = \frac{c^{1-\alpha}}{1 - \alpha}, \quad 0 < \alpha < \infty, \tag{3.2}$$

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

---

[6] See Constantinides et al. (2002, 2005). Constantinides et al. (2007) explicitly model bequests.

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. There is no fundamental economic reason why this must be so.

We assume 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 (2005), Danthine and Donaldson (2005), Duffie (2001), and LeRoy and Werner (2001).

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

For equity we have

$$1 = \beta E_t \left\{ \frac{U'(c_{t+1})}{U'(c_t)} R_{e,t+1} \right\}, \tag{3.3}$$

where

$$R_{e,t+1} = \frac{p_{t+1} + y_{t+1}}{p_t} \tag{3.4}$$

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

$$1 = \beta E_t \left\{ \frac{U'(c_{t+1})}{U'(c_t)} \right\} R_{f,t+1}, \tag{3.5}$$

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

$$R_{f,t+1} = \frac{1}{q_t} \tag{3.6}$$

with q_t being the price of the bond. Since $U(c)$ is assumed to be increasing we can rewrite (3.3) as

$$1 = \beta E_t \{ M_{t+1} R_{e,t+1} \}, \tag{3.7}$$

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 will hold. A little algebra shows that

$$E_t(R_{e,t+1}) = R_{f,t+1} + \text{Cov}_t \left\{ \frac{-U'(c_{t+1}), R_{e,t+1}}{E_t(U'(c_{t+1}))} \right\}. \tag{3.8}$$

The equity premium $E_t(R_{e,t+1}) - R_{f,t+1}$ can thus be easily computed. Expected asset returns equal the risk-free rate plus a premium for bearing risk, which depends on the covariance of the asset returns with the marginal utility of consumption. Assets that co-vary positively with consumption — that is, they pay off 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 US equity markets?

To address this issue, we make some additional assumptions. While they are not necessary and were not, in fact, part of our original paper on the equity premium, we include them to facilitate exposition and because they result in closed form solutions.[7]

---

[7] The exposition below is based on Abel (1988), his unpublished notes and Mehra (2003).

These assumptions are:

(a) the growth rate of consumption $x_{t+1} \equiv \frac{c_{t+1}}{c_t}$ is i.i.d.
(b) the growth rate of dividends $z_{t+1} \equiv \frac{y_{t+1}}{y_t}$ is i.i.d.
(c) (x_t, z_t) are jointly log-normally distributed.

The consequences of these assumptions are that the gross return on equity $R_{e,t}$ (defined above) is i.i.d, and that $(x_t, R_{e,t})$ are jointly log-normal.

Substituting $U'(c_t) = c_t^{-\alpha}$ in the fundamental pricing relation[8]

$$p_t = \beta E_t\{(p_{t+1} + y_{t+1})U'(c_{t+1})/U'(c_t)\} \tag{3.9}$$

we get

$$p_t = \beta E_t\{(p_{t+1} + y_{t+1})x_{t+1}^{-\alpha}\}. \tag{3.10}$$

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

$$p_t = w y_t$$

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

$$R_{e,t+1} = \frac{(w+1)}{w} \cdot \frac{y_{t+1}}{y_t} = \frac{w+1}{w} \cdot z_{t+1} \tag{3.11}$$

It is easily shown that

$$w = \frac{\beta E_t\{z_{t+1}x_{t+1}^{-\alpha}\}}{1 - \beta E_t\{z_{t+1}x_{t+1}^{-\alpha}\}} \tag{3.12}$$

hence

$$E_t\{R_{e,t+1}\} = \frac{E_t\{z_{t+1}\}}{\beta E_t\{z_{t+1}x_{t+1}^{-\alpha}\}}. \tag{3.13}$$

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

$$R_{f,t+1} = \frac{1}{\beta} \frac{1}{E_t\{x_{t+1}^{-\alpha}\}}. \tag{3.14}$$

---

[8] In contrast to our approach, which is in the applied general equilibrium tradition, there is another tradition of testing Euler equations (such as Equation (3.9)) and rejecting them. Hansen and Singleton (1982) and Grossman and Shiller (1981) exemplify this approach.

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

$$E_t\{R_{e,t+1}\} = \frac{e^{\mu_z + \frac{1}{2}\sigma_z^2}}{\beta e^{\mu_z - \alpha\mu_x + \frac{1}{2}(\sigma_z^2 + \alpha^2\sigma_x^2 - 2\alpha\sigma_{x,z})}} \tag{3.15}$$

and

$$\ln E_t\{R_{e,t+1}\} = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 + \alpha\sigma_{x,z}, \tag{3.16}$$

where $\mu_x = E(\ln x)$, $\sigma_x^2 = \text{Var}(\ln x)$, $\sigma_{x,z} = \text{Cov}(\ln x, \ln z)$, and $\ln x$ is the continuously compounded growth rate of consumption. The other terms involving z and R_e are defined analogously.

Similarly

$$R_f = \frac{1}{\beta e^{-\alpha\mu_x + \frac{1}{2}\alpha^2\sigma_x^2}} \tag{3.17}$$

and

$$\ln R_f = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 \tag{3.18}$$

Therefore,

$$\ln E\{R_e\} - \ln R_f = \alpha\sigma_{x,z}. \tag{3.19}$$

From (3.11) it also follows that

$$\ln E\{R_e\} - \ln R_f = \alpha\sigma_{x,R_e}, \tag{3.20}$$

where

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

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

$$\ln E\{R_e\} - \ln R_f = \alpha\sigma_x^2 \tag{3.21}$$

and the equity premium then is the product of the coefficient of relative risk aversion and the variance of the growth rate of consumption. As we see below, this variance is 0.001369, 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 and Prescott (1985) we reported the following sample statistics for the US economy over the period 1889–1978:

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 were 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.[9] 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 premia using the parameterization above?

Using the expressions derived earlier, we have

$$\ln R_f = -\ln \beta + \alpha \mu_x - \frac{1}{2}\alpha^2 \sigma_x^2 = 0.124$$

or

$$R_f = 1.132$$

that is, a risk-free rate of 13.2 percent!

Since

$$\ln E\{R_e\} = \ln R_f + \alpha \sigma_x^2 = 0.136$$

we have

$$E\{R_e\} = 1.146$$

---

[9] A number of these studies are documented in Mehra and Prescott (1985).

or a return on equity of 14.6 percent. This implies an equity risk premium of 1.4 percent, far lower than the 6.18 percent historically observed equity premium. In this calculation, we have been liberal in choosing 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 risk-free rate will be even higher and the premium lower. So the 1.4 percent 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 percent, we have a puzzle on our hands that risk considerations alone cannot account for.

3.2   The Risk-Free Rate Puzzle

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 (3.18) can be decomposed into three components.

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

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 greater 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, $\frac{1}{2}\alpha^2 \sigma_x^2$, 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"
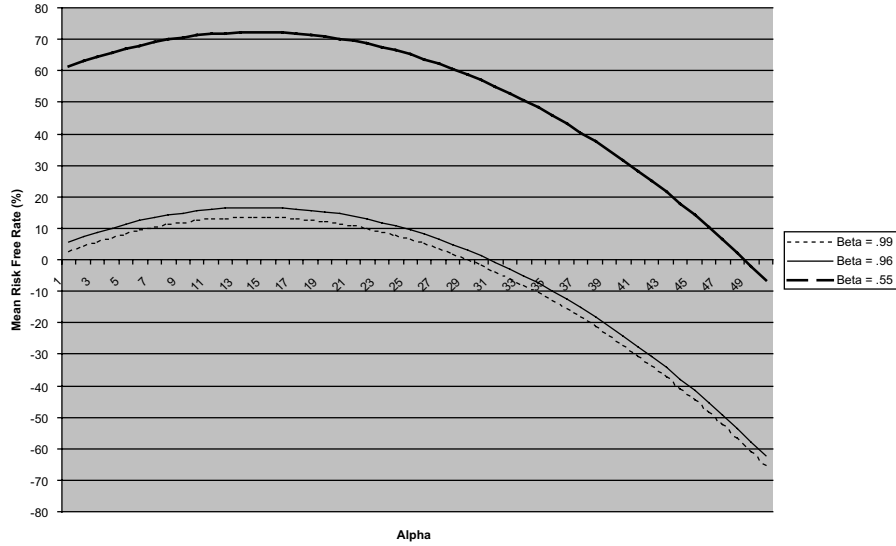
Fig. 3.3 Mean risk-free rate vs. alpha.

of the consumption good. Hence, in equilibrium the interest rate must fall to counter this enhanced demand for savings.

Figure 3.3, plots $\ln R_f = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2$ calibrated to the US 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 $(1 < \alpha < 5)$.

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[10] proposed that $\alpha = 55$ would solve the puzzle. Indeed it can be shown that the 1889–1978 US experience reported above can be reconciled with $\alpha = 48$ and $\beta = 0.55$.

---

[10] Private communication, 1981.

To see this, observe that since

$$\sigma_x^2 = \ln\left[1 + \frac{\text{var}(x)}{[E(x)]^2}\right] = 0.00123$$

and

$$\mu_x = \ln E(x) - \frac{1}{2}\sigma_x^2 = 0.0175$$

this implies

$$\alpha = \frac{\ln E(R) - \ln R_F}{\sigma_x^2} = 48.4.$$

Since

$$\ln\beta = -\ln R_F + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 = -0.60$$

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 3.3. An alternate approach is to experiment with negative time preferences; however there seems to be no empirical evidence that agents do have such preferences.[11]

Figure 3.3 shows that for extremely high α the precautionary savings term dominates and results in a "low" risk-free rate.[12] However, then a small change in the growth rate of consumption will have a large impact on interest rates. This is inconsistent with a cross-country comparison of real risk-free rates and their observed variability. For example, throughout the 1980s, South Korea had a much higher growth rate than the United States but real rates were not appreciably higher. Nor does the risk-free rate vary considerably over time, as would be expected if α was large. In Section 4, we show how alternative preference structures can help resolve the risk-free rate puzzle.

---

[11] In a model with growth, equilibrium can exist with $\beta > 1$. See Mehra (1988) for the restrictions on the parameters α and β for equilibrium to exist.

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

3.3   The Effect of Serial Correlation in the Growth Rate of Consumption

The analysis above has assumed that the growth rate of consumption is i.i.d over time. However, for the sample period 1889–2004 it is slightly negative (-0.1353) while for the sample period 1930–2004 the value is 0.45. The effect of this non zero serial correlation on the equity premium can be analyzed using the framework in Mehra and Prescott (1985). Figure 3.4 shows the effect of changes in the risk aversion parameter on the equity premium for different serial correlations. When the serial correlation of consumption is positive, the equity premium actually declines with increasing risk aversion, further exacerbating the equity premium puzzle.[13]
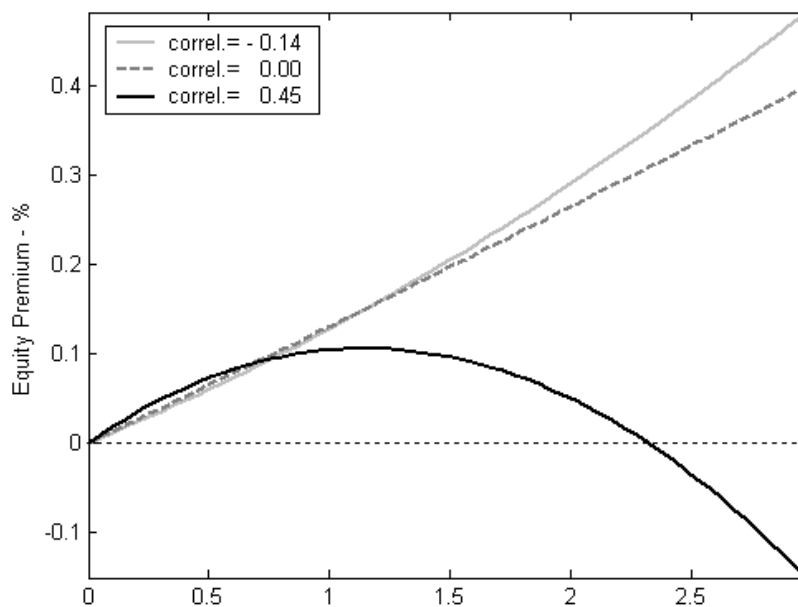

Fig. 3.4 The effect of changes in the risk aversion parameter on the equity premium fir different serial correlations.

---

[13] See Azeredo (2007) for a detailed discussion.

3.4   Hansen–Jagannathan Bounds

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

$$E_t(R_{e,t+1}) = R_{f,t+1} - \text{Cov}_t\left\{\frac{M_{t+1}, R_{e,t+1}}{E_t(M_{t+1})}\right\}. \qquad (3.22)$$

This expression also holds unconditionally so that

$$E(R_{e,t+1}) = R_{f,t+1} - \sigma(M_{t+1})\sigma(R_{e,t+1})\rho_{R,M}/E_t(M_{t+1}) \qquad (3.23)$$

or

$$E(R_{e,t+1}) - R_{f,t+1}/\sigma(R_{e,t+1}) = -\sigma(M_{t+1})\rho_{R,M}/E_t(M_{t+1}) \qquad (3.24)$$

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

$$|E(R_{e,t+1}) - R_{f,t+1}/\sigma(R_{e,t+1})| \leq \sigma(M_{t+1})/E(M_{t+1}). \qquad (3.25)$$

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

For the US 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 $\alpha = 2$. This implies 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 20 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,[14] modified probability distributions to admit rare but disastrous events,[15] survival bias,[16] incomplete markets,[17] and market imperfections.[18] They also include attempts at modeling limited participation of consumers in the stock market,[19] and problems of temporal aggregation.[20] We examine some of the research efforts to resolve the puzzle[21] in Sections 4 and 5.

Section 4 addresses risk and preference based explanations of the equity premium. Section 5, in contrast, reviews the nascent literature that takes as given the findings in Mehra and Prescott (1985) and tries to account for the equity premium by *factors other than aggregate risk*.

---

[14] For example, Abel (1990), Bansal and Yaron (2004), Benartzi and Thaler (1995), Boldrin et al. (2001), Campbell and Cochrane (1999), Constantinides (1990), Epstein and Zin (1991), and Ferson and Constantinides (1991).

[15] See, Rietz (1988) and Mehra and Prescott (1988).

[16] See Brown et al. (1995)

[17] For example, Bewley (1982), Brav et al. (2002), Constantinides and Duffie (1996), Heaton and Lucas (1997, 2000), Lucas (1994), Mankiw (1986), Mehra and Prescott (1985), Storesletten et al. (2007), and Telmer (1993).

[18] For example, Aiyagari and Gertler (1991), Alvarez and Jermann (2000), Bansal and Coleman (1996), Basak and Cuoco (1998), Constantinides et al. (2002), Danthine et al. (1992), Daniel and Marshall (1997), He and Modest (1995), Heaton and Lucas (1996), Luttmer (1996), McGrattan and Prescott (2001), and Storesletten et al. (2004).

[19] Attanasio et al. (2002), Brav et al. (2002), Mankiw and Zeldes (1991), and Vissing-Jorgensen (2002).

[20] Gabaix and Laibson (2001), Heaton (1995), Lynch (1996).

[21] The reader is also referred to the excellent surveys by Narayana Kocherlakota (1996), John Cochrane (1997) and by John Campbell (1999, 2001).

4

Risk Based Explanations of the Equity Premium

4.1 Alternative Preference Structures

4.1.1 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

$$U_t = \{(1-\beta)c_t^\rho + \beta\{E_t(\tilde{U}_{t+1}^\alpha)\}^{\frac{\rho}{\alpha}}\}^{\frac{1}{\rho}}, \tag{4.1}$$

where $1-\alpha$ is the coefficient of relative risk aversion and $\sigma = \frac{1}{1-\rho}$ is 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[1]

$$k_{t+1} = \beta^{\frac{\alpha}{\rho}} \left(\frac{c_{t+1}}{c_t}\right)^{\frac{\alpha(\rho-1)}{\rho}} (1 + R_{w,t+1})^{\frac{\alpha-\rho}{\rho}}. \tag{4.2}$$

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

$$p_t = E_t(k_{t+1}y_{t+1}). \tag{4.3}$$

In this framework the asset is priced *both* by its covariance with the growth rate of consumption (the first term in 4.2) *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, $1-\alpha$, 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 (3.3) and (3.5) 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

---

[1] Epstein and Zin (1991) use dynamic programming to calculate this. See their Equations 8–13. Although the final result is correct, there appear to be errors in the intermediate steps.

the "market portfolio" 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 3, and using the market portfolio as a stand in for the wealth portfolio we have

$$\ln R_f = -\ln \beta + \frac{\mu_x}{\sigma} - \frac{\alpha/\rho}{2\sigma^2}\sigma_x^2 + \frac{(\alpha/\rho) - 1}{2}\sigma_m^2. \qquad (4.4)$$

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.

4.2 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." Internal habit formation captures the notion that utility is a decreasing function of *one's 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 one's 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

$$U(c) = E_t \sum_{s=0}^{\infty} \beta^s \frac{(c_{t+s} - \lambda c_{t+s-1})^{1-\alpha}}{1-\alpha}, \quad \lambda > 0. \qquad (4.5)$$

If $\lambda = 1$ 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.[2] The implied local coefficient of relative risk aversion is

$$-\frac{cu''}{u'} = \frac{\alpha}{1 - x/c}. \qquad (4.6)$$

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 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.[3]

An alternate approach to circumvent this problem has been expounded by Campbell and Cochrane (1999). Their model incorporates the possibility of recession as a state variable so that risk aversion varies in a highly nonlinear manner.[4] 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

---

[2] See also the discussion in Weil (1989).

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

[4] If we linearize the "surplus consumption ratio" in the Campbell and 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.

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[5] aggregate consumption. Abel (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 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.[6] 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

$$U(c) = E_t \sum_{t=0}^{\infty} \beta^t \frac{(c_t/C_{t-1}^{\gamma})^{1-\alpha}}{1-\alpha}, \quad \alpha > 0, \tag{4.7}$$

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

---

[5] Hence "Catching up with the Joneses" rather than "keeping up with the Joneses." Abel (1990) footnote 1.

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

Equations (3.3) and (3.5) below.

$$1 = \beta E_t\{R_{e,t+1}x_t^{\gamma(\alpha-1)}x_{t+1}^{-\alpha}\} \tag{4.8}$$

$$1 = \beta R_{f,t+1}E_t\{x_t^{\gamma(\alpha-1)}x_{t+1}^{-\alpha}\}, \tag{4.9}$$

where $x_{t+1} \equiv \frac{c_{t+1}}{c_t}$ is the growth rate of consumption. Under the assumptions made in Section 3.1 we can write

$$R_{f,t+1} = E_t\{x_{t+1}^{\gamma(\alpha-1)}\}/\beta E_t\{x_{t+1}^{-\alpha}\} \tag{4.10}$$

and

$$E_t\{R_{e,t+1}\} = E_t\{x_{t+1}^{\gamma(\alpha-1)}\}\left(\left[E_t\{x_{t+1}\} + AE_t\{x_{t+1}^{1+\gamma(\alpha-1)}\}\right]/A\right). \tag{4.11}$$

We see that in the expression $\ln R_f = -\ln\beta + \alpha\mu_x - \frac{1}{2}\alpha^2\sigma_x^2 - \gamma(1-\alpha)\mu_x$, 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).

4.3   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 $1 + R_{i,t+1}$.

Abel (1990) reports expressions for $E(1 + R_{i,t+1})$ and $E(1 + R_{f,t+1})$ in Equations 17 and 18.

Let $\prod_{\text{Abel}} = \ln(E(1 + R_{i,t+1})) - \ln(E(1 + R_{f,t+1}))$. In the Abel model with $\theta = 0$ (the "standard time additive" case), *if* the growth rate of consumption is assumed to be log-normally distributed, $\prod_{\text{Abel}}$ can be written as:

$$\prod_{\text{Abel}} = E(\ln(1 + R_{i,t+1})) + 0.5\text{Var}(\ln(1 + R_{i,t+1}))$$
$$- E(\ln(1 + R_{f,t+1})) - 0.5\text{Var}(\ln(1 + R_{f,t+1})) \quad (4.12)$$

or

$$\prod_{\text{Abel}} = \prod_{\text{Campbell}} + 0.5[\text{Var}(\ln(1 + R_{i,t+1})) - \text{Var}(\ln(1 + R_{f,t+1}))]$$
$$(4.13)$$

or

$$\prod_{\text{Abel}} = \prod_{\text{Campbell}} + 0.5\text{Var}(\ln(x)), \quad (4.14)$$

where $\prod_{\text{Campbell}} = E(\ln(1 + R_{i,t+1})) - E(\ln(1 + R_{f,t+1}))$, 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 $\prod_{\text{Abel}}$ from Equations 17 and 18 in Abel (1990) yields $\prod_{\text{Abel}} = \alpha\text{Cov}(\ln x, \ln(1 + R_i))$. This is identical to that obtained by adjusting Equation 62 in Campbell by adding $0.5\text{Var}(\ln(x))$.

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).

4.4   The Campbell and Cochrane (1999) Mechanism

Campbell and Cochrane (1999) postulate a process on an external habit, X_t, and a period utility function of the simple CRRA form:

$$u(C_t, X_t) = \frac{(C_t - S_t)^{1-\gamma}}{1 - \gamma}, \tag{4.15}$$

where C_t in this case denotes the individual agent's consumption and $S_t = \left(\frac{C_t - X_t}{C_t}\right)$ his surplus consumption habit. Note that from the agent's perspective, S_t is exogenous although in equilibrium this will not be so. They next postulate

(i) consumption growth follows an i.i.d. log-normal process,

$$\Delta c_{t+1} = \log C_{t+1} - \log C_t \equiv g + \tilde{\nu}_t$$
$$\text{where } \tilde{\nu}_t \sim \text{ i.i.d. } N(0, \sigma_\nu^2) \text{ and} \tag{4.16}$$

(ii) the log of the surplus consumption ratio $\log S_t = s_t$ also follows a log-normal process to be consistent with (i)

$$s_{t+1} = (1 - \phi)\bar{s} + \phi s_t + \lambda(s_t)(\Delta c_{t+1} - g), \tag{4.17}$$

where ϕ, $\bar{s}$, and g are parameters, $\bar{s} = \log \bar{S}$ ($\bar{S}$ denotes the steady state surplus consumption ratio), and $\lambda(s_t)$ is a pre-specified sensitivity function. As a result, the period $t, t+1$ pricing kernel is of the form:

$$\begin{aligned}
\text{MRS}_{t,t+1} &= \beta \frac{u_1(C_{t+1}, S_{t+1})}{U_1(C_t, S_t)} = \beta \left(\frac{S_{t+1}}{S_t} \cdot \frac{C_{t+1}}{C_t}\right)^{-\gamma} \\
&= \beta e^{-\gamma[(1-\phi)(\bar{s}-s_t)+(1+\lambda(s_t))\Delta c_{t+1}-\lambda(s_t)g]} \\
&= \beta e^{-\gamma[(\phi-1)(s_t-\bar{s})+[1+\lambda(s_t)]\tilde{\nu}_{t+1}]} \tag{4.18}
\end{aligned}$$

by identification (i) above. Under this specification, it is apparent that the standard deviation of the MRS, as well as its correlation with consumption growth (both important determinants of the premium) will be determined by the form of the sensitivity function. Campbell and Cochrane (1999) go

on to make the following assumption on the form of $\lambda(s_t)$,

$$\lambda(s_t) = \begin{cases} \frac{1}{\bar{S}}\sqrt{1 - 2(s_t - \bar{s})} - 1 & s_t \leq s_{\max} \\ 0 & s_t \geq s_{\max}, \end{cases} \quad (4.19)$$

where $\bar{S} = \sigma\sqrt{\beta/(1 - \phi)}$, and $s_{\max} = \bar{s} + (1/2)(1 - \bar{S}^2)$.

These requirements allow the habit formulation not only to reflect certain key features of the data but also to generate a plausible habit process. In doing so the authors attempt to set an upper bound on what can be achieved with a habit model in an exchange setting; that is, they seek to provide a sort of "performance standard" that other competing habit formulations must exceed. No axiomatic foundations for the postulated surplus consumption process are proposed, however.

Finally, an i.i.d. process on dividend growth is postulated, one that is imperfectly correlated with growth in consumption:

$$\Delta d_{t+1} = g + \tilde{\omega}_{t+1}, \quad \tilde{\omega}_{t+1} \sim \text{ i.i.d. } N(0, \sigma_\omega^2) \text{ with } \text{corr}(\tilde{\omega}_t, \tilde{\nu}_t) = \rho. \quad (4.20)$$

As before, d_t denotes the log of the actual period t dividend, D_t. With an independently specified dividend process, the $C_t = Y_t$ identity characteristic of Mehra and Prescott (1985) is broken, allowing an implicit wage process, levered capital structure, etc. The parameters ρ *and* σ_ω remain to be specified.

The model — and, in particular, the surplus consumption process — is designed to replicate the observed low risk-free rate volatility. Taking advantage of the well-known properties of the log-normal return process in conjunction with model specifications (4.15)–(4.18), it becomes a matter of straightforward manipulation to derive the implied expression for log r_t^f:

$$\ln r_t^f = -\ln\beta + \gamma g - \gamma(1 - \phi)(s_t - \bar{s}) - \frac{\gamma^2\sigma^2}{2}(1 + \lambda(s_t))^2. \quad (4.21)$$

Substituting Equation (4.17), Equation (4.21) can be written as

$$\ln r_t^f = -\ln\beta + \gamma_g - \frac{\gamma}{2}(1 - \phi), \quad (4.22)$$

a constant. This clearly is an extreme case of low volatility, and Campbell and Cochrane (1999) offer generalizations by which the precise real risk-free rate volatility can be matched as well.

Plausibility dictates that the habit should not be predetermined, except possibly at the steady state, and formulation (4.17), (4.19) also satisfies this requirement. Indeed the fact that Δc_{t+1} appears in the expression for the period $t+1$ surplus consumption ratio s_{t+1} dictates that current consumption growth has some influence on the habit.[7] Since the habit is external, it represents the influence on an individual's preference ordering of the average consumption plan of others in the society. Even in an environment of instant communication, it is perhaps unreasonable to presume an immediate and fully external habit adjustment process.

The sensitivity parameter $\lambda(s_t)$ is also increasing in s_t by construction. This guarantees countercyclical risk aversion and, as the authors demonstrate, a countercyclical market price of risk. By making the agent very risk averse when consumption is already low, this feature accentuates the perceived riskiness of a variable consumption stream. Table 4.1 presents some sample results of this elaborate exercise.

From these results it is clear that habit formation as articulated by Campbell and Cochrane (1999) is a powerful mechanism for influencing

Table 4.1 Equilibrium security returns and the premium in a Campbell–Cochrane model rates of return, growth rates annualized in percent.

	Postwar data	Simulated moments
Er^f	0.94	0.094
σr^f	(not reported)	0
Er^e	7.58	7.63
σr^e	(not reported)	15.2
Er^p	6.69	6.64
σr^p	15.7	15.2
$E\Delta c_t$	1.89	1.89
$\sigma\Delta c_t$	1.22	1.22

Notes: $g = 1.89$, $\sigma = 1.50$, $\ell n r^f = 0.94$, $\phi = 0.87$, $\rho = 0.20$, $\sigma_\omega = 11.2$, $\beta = 0.89$, $\bar{S} = 0.057$, $S_{\max} = 0.094$.
Source: Campbell and Cochrane (1999), various tables.

---

[7] Including current consumption as a part of the habit formation process may be used to ensure that period t's consumption never falls below period t's habit. Otherwise period utility may not be well defined.

security returns in a dynamic stochastic general equilibrium model. Its flexibility is even greater than what Table 4.1 suggests since Campbell and Cochrane (1999) subject the model to other tests (e.g., predictability) which it easily transcends. Yet we remain uneasy, as there is no empirical evidence to support the assertion that surplus consumption habits evolve in the way proposed. Furthermore, for the reported simulation, the agent's effective CRRA varies between one and one hundred, which is arguably extreme.

The surplus consumption habit process in Campbell and Cochrane (1999) is highly specialized in other ways. Ljungqvist and Uhlig (1999), in particular, point out that under their specification, the representative agent would experience substantial welfare gains if 10 percent of his endowment were periodically destroyed. The basic intuition is straightforward: while utility is diminished in the period in which consumption endowment is destroyed, future utility gains result since the habit is correspondingly lower. If the former loss is more than compensated by the latter gains, the overall result is welfare enhancing. While this is never the case under standard linear habit evolution, it is possible under the Campbell and Cochrane (1999) construct.[8] These anomalies suggest the need for an axiomatic treatment of habit formation.

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.

---

[8] These observations are not as general as would initially appear. Consider the representative agent utility function $\sum \beta^t \{g(x_t) + v(c_t, x_t)\}$, where x_t is the aggregate consumption history, c_t is the agent's time-t consumption and $v(c_t, x_t)$ is increasing and concave in c_t; $g(x_t)$ and $v(c_t, x_t)$ together constitute the agent's period utility function. With an external habit, marginal utility is given by $\partial v(c_t, x_t)/\partial c_t$, independent of $g(x_t)$. The class of utility functions with common $v(c_t, x_t)$ but different $g(x_t)$ support the same equilibrium but may, in general, have different welfare implications. Campbell and Cochrane indeed focus on equilibrium implications driven exclusively by marginal utility $\partial v(c_t, x_t)/\partial c_t$. It is not the case that all utility functions in the class will necessarily exhibit identical welfare implications and therefore the criticism leveled by Ljungquist and Uhlig (1999) against Campbell–Cochrane cannot be viewed as a general statement. We thank George Constantinides for pointing this out to us.

4.5 Idiosyncratic and Uninsurable Income Risk

At a theoretical level, aggregate consumption is a meaningful economic construct if the market is complete, or effectively so.[9] 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.[10]

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.[11]

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

---

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

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

[11] 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.

instance, consumers face the risk of job loss, or other major personal disasters that cannot be hedged away or insured against.[12] 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*.

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 et al. (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

---

[12] Storesletten et al. (2004, 2007) provide empirical evidence from the Panel Study on Income Dynamics (PSID) that idiosyncratic income shocks are persistent and have counter cyclical conditional variance and corroborate this evidence by studying household consumption over the life cycle.

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.[13] 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.

4.6   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 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 100 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 had risen significantly as the war years receded, that would have provided evidence in support of the Rietz hypothesis. Similarly, if the low

---

[13] 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 (2004) study the CEX database by constructing synthetic cohorts; and Ait-Sahalia et al. (2001) measure household consumption by the purchases of certain luxury goods.

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.

Barro (2006) presents a model in the same spirit as Rietz (1988) but in a context where more attention may be given to calibration and the historical record. The historical justification for his calibration is, in fact, fascinating to read. The point of departure of the paper is the postulation of a stochastic process for aggregate output growth which allows for rare events in a generalized Rietz setting. Specifically, he models aggregate output growth as a random walk with drift, whose innovations are of three possible types:

(a) "diffusive" shocks (i.i.d. normal shocks),
(b) jump shocks of "type v" — shocks which represent situations where output contracts sharply but there is no occurrence of a default on debt, and
(c) jump shocks of "type w" — shocks which represent situations where output contracts sharply and a default on debt ensues.[14]

In particular, Barro assumes that the log of output $Y_t \equiv C_t$ grows according to the random walk with drift process

$$\log Y_{t+1} = \log Y_t + \bar{g} + \tilde{u}_{t+1} + \tilde{v}_{t+1}, \qquad (4.23)$$

where $\tilde{u}_{t+1}$ is an i.i.d random variable distributed $N(0,\sigma_u^2)$ and $\tilde{v}_{t+1}$ (jump shock "type v") captures low probability downward jumps in GDP. If a disaster occurs, with probability p, output declines by the fraction b. Very roughly speaking, a reasonable calibration requires b to be large and p small.

---

[14] Barro (2006) does not explicitly define a type w shock. We introduce it here as an aid to exposition.

More formally,

$$\tilde{v}_{t+1} = \begin{cases} 0, & \text{with probability } e^{-p} \\ \log(1 - \tilde{b}), & \text{with probability } 1 - e^{-p}. \end{cases} \quad (4.24)$$

In his calibration exercise, $\tilde{b}$ is a random variable whose probability density function coincides with the frequency distribution reflecting the size of contractions in 35 countries in the 20th century ($\tilde{b}$ ranges from 0.17 to 0.62). Barro also admits the possibility of default on government debt of proportion d, with probability q whenever a "type v" jump occurs. In other words, whenever a jump shock of "type v" occurs there is a probability q of a "type w" jump. Table 4.2 presents a sample of results for the calibration exercises ($\tilde{b} = \tilde{d}$). In all cases, preferences are as in Mehra and Prescott (1985) with CRRA $= 4$ and time preferences parameter $\beta = e^{-\rho} = e^{-0.03} = 0.9704$.

The results presented here are intriguing and lend support to a resurrection of the disaster scenario as a viable justification for the premium.

The distinction between "type v" events and "type w" events is theoretically appealing in terms of their different implications for quantities and prices. However, we feel this distinction is largely a consequence of the specific functional forms used to model the two types of rare events and the assumption that they are independent–something we intuitively think is not very plausible.

To support the observed equity premium (in the baseline model) a decline in GDP growth of 50 percent is needed. The plausibility of such a substantial decline in per capita GDP growth is justified by the figures in Barro (2006), Table 1 where this decline is shown to be 64 percent in Germany and Greece (during World War II) or 31 percent in

Table 4.2 Financial statistics: Barro's (2006) model (all rates of return annualized in percent).

	$\bar{g} = 0.025$ $p = 0.017$	$\bar{g} = 0.025$ $p = 0.025$	$\bar{g} = 0.020$ $p = 0.017$	$\bar{g} = 0.025$ $p = 0.017$
Er^e	7.1	4.4	5.1	6.1
Er^f	3.5	−0.7	1.5	2.5
Er^p	3.6	5.2	3.6	3.6

Notes: Data used for calculation: All cases $\theta = 4$, $\sigma_u = 0.02$, $q = 0.4$, Time preference $e^{-\rho}$, $\rho = 0.03$.

the United States (during the Great Depression). These observations however, are not at an annual frequency.[15] Looking at data at an annual frequency, we see little evidence of such a substantial decline either in GDP growth or in the growth rate of consumption. If the equity premium is estimated at an annual frequency, then the relevant growth rates should also be at annual frequencies. We argue that calibrations should be done at the same frequency as observations.

As is standard in the equity premium literature, Barro calibrates his model using output growth rates rather than consumption growth rates since in a pure exchange setting they are identical. While this distinction may not matter in a stylized economy without disaster states, the two quantities might empirically be quite different in the event of an extreme disaster. Consumption per capita will probably be a much smoother series than GDP per capita. In the event that such states occur, consumption is unlikely to remain a constant proportion of the output. Capital investment and capital utilization rates will change and act as channels to absorb the negative productivity shocks (see also Section 3) enabling consumption smoothing over time and across states. For this reason, since we do live in a production economy, it would be interesting to explore the robustness of Barro's (2006) results using cross-country data on the growth rates of per capita consumption instead of output within that more general context.

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:

---

[15] In a sense the results are obtained using the four-year drop in consumption as if it is a one-year effect. The claimed equity premium should be compared to the observed four-year effect $4 \times 6 = 24$ percent and not the one-year (6 percent).

During the German hyperinflation, holders of bonds denominated in Reich marks lost virtually all the value invested in those assets. During the Poincaré administration in France during the 1920s, bond holders lost nearly 90 percent 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.

4.7   Behavioral Models

Behavioral models are ones for which the postulated agents are not "fully rational." While this can mean that they do not have rational expectations[16] or that they do not update their beliefs using Bayes' rule, in the context of this paper it will be taken to mean that the agents make decisions in accordance with preferences which are not expected utility defined over consumption or wealth (indirect utility). In this sense "habit formation" preferences are behavioral, although their successful use in explaining a wide class of phenomena has given them a degree of acceptance not typical of the broader behavioral finance literature. While the proposed family of behavioral preference structures

---

[16] By this we mean, in the context of this paper, that agents know the exact probability distribution governing all random outcomes — returns, consumption, etc. — driving the model.

is otherwise quite large, most have not been applied to the study of the "equity premium" and associated puzzles.[17] For this reason we restrict our attention to (1) "Disappointment Aversion" and "Generalized Disappointment Aversion" of Gul (1991) and Routledge and Zin (2004) and (2) to the "Prospect Theory" of Kahneman and Tervsky (1979) as applied, most especially, by Barberis et al. (2001). A more extensive discussion can be found in Donaldson and Mehra (2008).

4.8 Generalized Disappointment Aversion

Disappointment averse preferences place a greater utility weight on outcomes that disappoint. For Gul (1991), this threshold is to be identified with consumption realizations that lie below the (conditional) certainty equivalent consumption level while for Routledge and Zin (2004) outcomes more distant from the certainty equivalent are given special weighting.[18]

As in the Epstein and Zin (1989) utility construct, time and risk preferences are separated. Their recursive intertemporal utility function $U(C_t, CE_{t+1})$ has the form:

$$U(C_t, CE_{t+1}) = \left[\left(1 - \frac{1}{1+\rho} \right) C_t^\gamma + \frac{1}{1+\rho} CE_{t+1}^\gamma \right]^{1/\gamma}, \qquad (4.25)$$

where CE_{t+1} is the certainty equivalent pertaining to random consumption in the following period. In functional form (4.25), ρ is the marginal rate of time preference and $\frac{1}{1-\gamma}$ the elasticity of intertemporal substitution.

Risk preferences and the certainty equivalent computation are specified as follows. Let us presume a pure Mehra and Prescott (1985) setting with a finite number of states and suppose in some period t, the output level is $Y_t^i = Y^i$. Temporal risk preferences are represented as a period utility function of the form $u(C) = \frac{C^\alpha}{\alpha} (\alpha \neq 0)$ or

---

[17] For the wider classes that have been used for macroeconomic studies, see Backus et al. (2004).

[18] Gul (1991) Disappointment Aversion (DA) preferences in a standard Mehra–Prescott (1985) economy do not per se resolve the premium. Bonomo and Garcia (1993) enhance the construct by incorporating a joint process on consumption and dividends that follow a Markov switching model — in addition to (DA) preferences. They are able to match the risk-free rate and the first two moments of the premium.

$u(C) = \log C (\alpha = 0)$. The period $t + 1$ certainty equivalent consumption level $CE_{t+1} = CE_{t+1}(\tilde{Y}^i)$ is then endogenously determined as the solution to the following definitional equation:

$$
\begin{aligned}
u(CE_{t+1}(\tilde{Y}^i)) &= \frac{(CE_{t+1}(Y^i))^\alpha}{\alpha} \\
&= \sum_j \pi_{ij} \frac{(Y^j)^\alpha}{\alpha} - \theta \\
&\quad \times \sum_{Y^j \le \delta CE_{t+1}(Y^i)} \pi_{ij} \left[\frac{(\delta CE_{t+1}(Y^i))^\alpha}{\alpha} - \frac{(Y^j)^\alpha}{\alpha} \right]
\end{aligned}
$$

(4.26)

If $\theta = 0$, structure (4.26) reduces to the definition of a certainty equivalent in a pure expected utility context.[19] Otherwise, there is a penalty, proportional to θ, for consumption realizations that lie sufficiently below the endogenously determined certainty equivalent as defined in (4.8). If $\delta = 1$, the certainty equivalent computation follows according to Gul (1991); if $\delta < 1$ the context is that of Routledge and Zin (2004). Note that there is no mention of any external reference point as in the (external) habit formation literature.

Nevertheless, the model generates counter-cyclical risk aversion, a feature which bodes well for its asset pricing characteristics. Furthermore, GDA preferences are homothetic, linear in probabilities, and have the same aggregation properties as CRRA expected utility structures. They admit consistent planning and Routledge and Zin (2004) provide the axiomatic underpinnings.

Table 4.3 gives a quantitative idea of how well GDA preferences perform in the standard Mehra and Prescott (1985) environment.

A number of these results are worthy of explicit mention. When dividend volatility is low, GDA preferences give only modest premia; they become quite large only when $\sigma(Y_t) = 0.10$. For the data period underlying the empirical background to the article, de-trended

---

[19] When $\theta \ne 0$, the Gul–Routledge–Zin structures relax the independence assumption that underlies the expected utility representation. Independence fails because mixing in some arbitrary payoff lottery with two lotteries under comparison may fundamentally change the tail properties of each.

Table 4.3 Equilibrium security returns and the premium: GDA preferences, various parameter combinations (all returns in percent).

	A	B	C	D	E	F
Parameter choices						
α	1	1	0.5	-2	1	1
γ	1	1	1	1	0.5	-0.5
θ	9	24	9	9	9	9
δ	0.9692	0.9692	0.9692	0.9692	0.9692	0.9692
ρ	0.03	0.03	0.03	0.03	0.03	0.03
$\sigma(Y_t) = 0.05$						
Er^f	1.74	1.56	1.72	1.62	2.45	3.89
σ_{r^e}	1.26	1.44	1.25	1.23	1.21	1.11
Er^e	4.30	4.68	4.33	4.46	5.07	6.63
σ_{r^e}	5.72	6.14	5.68	5.52	5.86	6.14
Er^p	2.56	3.12	2.61	3.84	3.86	2.74
σ_{r^p}			not reported			
$\sigma(Y_t) = 0.10$						
Er^f	1.74	1.56	1.72	1.62	2.45	3.89
σ_{r^e}	1.26	1.44	1.25	1.23	1.21	1.11
Er^e	8.07	9.66	8.17	8.59	8.86	10.46
σ_{r^e}	14.37	16.26	14.27	13.80	14.55	14.91
Er^p	6.33	8.10	6.45	6.97	6.65	9.35
σ_{r^p}			not reported			

Source: Table 4.3 is an amalgamation of selected results reported in Tables 1 and 2 in Routledge and Zin (2004).
Note: The return statistics for the risk-free rate are identical for the $\sigma(Y_t) = 0.05$ and 0.10 cases because risk-free securities do not result in disappointments.

dividend volatility is on the order of 12 percent which advocates for the model with $\sigma(Y_t) = 0.10$; Campbell (1999) argues that the standard deviation of the dividend should be as high as 28 percent. Qualitative results generally match intuition; comparing columns (A) and (B), the result of a greater penalty is to raise risk-free asset prices and lower equity ones: risky securities are less desirable and the premium rises. Comparing columns (B) and (D), the representative agent becomes atemporally more risk averse and the premium naturally increases. In general, equity volatilities ($\sigma(Y_t) = 0.10$ case) match the data quite well; risk-free rate volatilities are generally too small. Although not part of Table 4.3, it can also be shown that the risk-free rate is procyclical. Overall, the results are very good, and the model is advantaged by retaining a utility specification defined only over the agent's consumption. Note

also that countercyclical risk aversion is generated endogenously within the (GDA) construct. To see this, consider the cases of columns A and B: disappointment aversion is the only source of risk aversion ($\alpha = 1$). As a result, the pricing kernel is risk neutral in the high growth states (no disappointing outcomes) but substantially risk averse in the low growth states (see Routledge and Zin (2004) for a more detailed analysis).

There is one other major structure that provides countercyclical risk aversion: prospect theory as applied to asset pricing.

4.9   Prospect Theories

The major reference is Barberis et al. (2001). See also the work by Barberis and Huang (2008). The idea is to postulate investors who derive utility not only from their period-by-period real consumption, but also over *equity* portfolio gains and losses relative to a plausible benchmark.[20] Roughly speaking, these investors max $U(c) + V(G/L)$ where G/L denotes gains or losses. That utility is defined over only equity gains/losses rather than over aggregate gains and losses to total wealth is an illustration of the notion of "narrow framing."

Furthermore, investors are more sensitive to losses than to gains (this is the essence of the notion of "loss aversion"), with the extent of the loss sensitivity depending on the agent's prior portfolio experience. Equity gains and losses are measured with respect to a benchmark, and if the investor's past experience has been to sustain losses relative to his benchmark, then he is modeled as being more acutely sensitive to further ones. If his recent experience has been one of equity portfolio gains, then the agent is modeled as being less sensitive to losses (provided they are not so severe as to negate past gains). These requirements lead to a kink in the agent's utility curve at gain = loss = 0. Roughly speaking, loss aversion as captured in Barberis et al. (2001) implies linear utility over gains and losses to the equity portion of an

---

[20] See also Benartzi and Thaler (1995). These authors postulated investors with loss averse preferences over variations in their financial wealth only. Without any direct connection to consumption it is impossible to ascertain how their model might describe the joint processes on equilibrium returns and consumption growth, for example.

investor's portfolio of the form:

$$V(X_{t+1}, S_t, z_t) = \begin{cases} X_{t+1} & X_{t+1} \geq 0 \\ \lambda(z_t)X_{t+1} & X_{t+1} < 0, \end{cases}$$

where

$$X_{t+1} = S_t R_{t+1}^e - R_t^f S_t.$$

Following our customary notation, R_{t+1}^e and R_t^f, designate, respectively, the gross rates of return on stock and risk-free bonds from period t to $t+1$ and S_t the value of the stock portion of the investor's portfolio. The latter is subscripted by t alone as the risk-free rate from period t to $t+1$ is solely determined by the risk-free bond's price in period t. Gains or losses are thus measured relative to what would have been earned had the stock portion S_t of his portfolio been invested in risk-free assets. The slope parameter $\lambda(z_t)$ measures the degree of loss sensitivity as dependent on a variable, z_t, that recalls past portfolio experience, where $z_t = 1$ captures a case of no prior gains or losses, $z_t < 1$ prior gains and $z_t > 1$ prior losses. A representative graphical portrayal is found in Figure 4.1.

More formally, in a discrete time asset pricing context, the representative agent undertakes to assemble portfolios of the stock market index with value S_t in conjunction with an aggregate valuation of risk-free
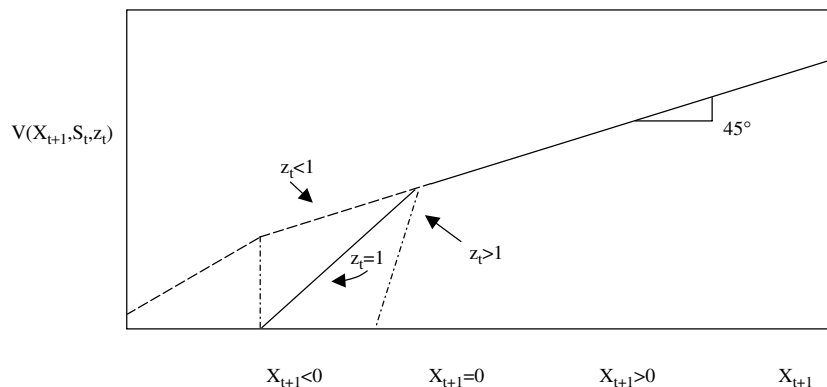

Fig. 4.1 $V(X_{t+1}, S_t, z_t)$: utility of equity portfolio gains or losses.
Source: From Barberis et al. (2001).

bonds, B_t, so as to solve:

$$\max E\left(\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma} + b_t \beta^{t+1} V(X_{t+1}, S_t, z_t)\right), \qquad (4.27)$$

where the first term is the standard discounted utility of consumption under CRRA utility, and the second, $V()$, captures the utility/disutility of gains and losses, X_{t+1}, in the stock market portion of the optimal portfolio, S_t; the z_t term is as indicated above. The standard period t budget constraint and the constraint governing the evolution of wealth are, respectively,

$$C_t + S_t + B_t \leq W_t \qquad (4.28)$$

$$W_{t+1} = S_t R_{t+1}^e + B_t R_t^f, \qquad (4.29)$$

where W_t represents period t wealth and S_t, B_t the amounts of it allocated to stock and risk-free bonds post consumption, C_t, in period t. It remains to detail the precise form of $V()$. Barberis et al. (2001) postulate

$$V(X_{t+1}, S_t, z)_t = \begin{cases} X_{t+1} & R_{t+1}^e \geq z_t R_t^f, \text{ and } z_t \leq 1 \\ S_t(z_t R_t^f - R_t^f) & R_{t+1}^e < z_t R_t^f, \text{ and } z_t \leq 1 \\ \quad + \lambda S_t(R_{t+1}^e - z_t R_t^f), & \\ X_{t+1} & R_{t+1}^e \geq R_t^f, \text{ and } z_t > 1 \\ \lambda(z_t) X_{t+1} & R_{t+1}^e < R_t^f, \text{ and } z_t > 1, \end{cases} \qquad (4.30)$$

where $\lambda(z_t)$ and z_{t+1} evolve according to

$$\lambda(z_t) = \lambda + k(z_t - 1), \qquad (4.31)$$

and

$$z_{t+1} = \eta\left(z_t \frac{\bar{R}^e}{R_{t+1}^e}\right) + (1-\eta)1, \qquad (4.32)$$

where $\lambda > 1$, $\eta > 0$, $k > 0$ are constants and z_t is an index which captures whether the investor has recently experienced gains ($z_t \leq 1$) or losses ($z_t > 1$).

Some careful interpretation is in order. Equation (4.32) describes the evolution of the historical experience variable, z_t, where $\bar{R}^e$ amounts

to the average stock return and η is a parameter which captures the memory in the adjustment process.[21] If equity returns are much less than average $(\bar{R}^e/R^e_{t+1} > 1)$, this experience tends to increase $z_{t+1} > z_t$ since $\eta \geq 0$. If $\eta = 0$, Equation (4.30) suggests that $z_t \approx 1$ or that prior gains and losses have an immediate effect on the investor's risk attitude. If $\eta = 1$ then the benchmark adjusts slowly. Note that $\lambda(z_t)$, the loss sensitivity, is increasing in z_t. Otherwise, $b_t = b_0 \bar{C}^{-\gamma}$, where $\bar{C}^{-\gamma}$ is purely a scaling factor which is included to maintain stationarity in the equilibrium asset return distributions and b_0 determines the relative utility of gains and losses vs. the utility of consumption. There is, however, no formal proof that investor preferences of this type survive aggregation and no axiomatic foundation for their existence.

It remains to close the economy by specifying the exogenous dividend and consumption processes (as in any exchange model there is one equity security outstanding, and risk-free bonds are priced in zero net supply). Barberis et al. (2001) consider a number of scenarios. We first report the one which is in the spirit of a pure Mehra and Prescott (1985) setting.

$$\log\left(\frac{C_{t+1}}{C_t}\right) = \log\left(\frac{D_{t+1}}{D_t}\right) = g_c + \sigma_c\tilde{\varepsilon}_{t+1}, \text{ where } \tilde{\varepsilon}_{t+1} \sim \text{ i.i.d. N}(1,\, 0),$$

where $g_c = 1.84$ and $\sigma_c = 3.79$ (both expressed in percent) as estimated from data for the period 1926–1995. The results are presented in Table 4.4.

Table 4.4 The equity premium in a model with loss aversion (annualized returns in percent).

		$b_0 = 0.7$ $k = 3$	$b_0 = 2$ $k = 3$	$b_0 = 100$ $k = 3$	$b_0 = 7$ $k = 150$	$b_0 = 2$ $k = 100$	$b_0 = 100$ $k = 50$
$\log r^f$		0.58	3.79	3.79	3.79	3.79	3.79
$E \log r^P$		6.03	0.63	0.88	1.26	3.50	3.66	3.28
$\sigma \log r^P$		20.02	4.77	5.17	5.62	10.43	10.22	9.35
Sharpe Ratio		0.3	0.13	0.17	0.22	0.34	0.36	0.35
Average Loss Aversion			2.25	2.25	2.25	10.7	7.5	4.4

Source: This is Table II of Barberis et al. (2001).
Notes: $\gamma = 1$, $\beta = 0.98$, $\lambda = 2.25$, $\eta = 0.9$, $g_c = 1.84\%$, $\sigma_c = 3.79\%$; Under their specification, it can be shown that $\log r_f$ is constant.

---

[21] $\bar{R}^e$ is actually determined within the model in order that, at equilibrium, the median value of $Z_t \equiv 1$. This turns out, not surprisingly, to be approximately the average equity return.

Note that for very large k — which leads via Equation (4.29) to very high loss aversion in some states of the world, the premium can rise to a level of 3.50 percent in conjunction with about half (10.4 percent) the observed historical equity return volatility. Increases in the value of the parameter b_0, which place greater overall relative weight on the utility of gains and losses also lend to increase the premium but only to 1.26 percent when $b_0 = 100$. Clearly, it is the loss aversion parameter that has the power.

Table 4.5 presents results when the consumption and dividend processes are specified independently:

$$\log\left(\frac{C_{t+1}}{C_t}\right) = g_c + \sigma_c \tilde{\xi}_{t+1}$$

$$\log\left(\frac{D_{t+1}}{D_t}\right) = g_D + \sigma_D \tilde{\varepsilon}_{t+1}$$

$$\begin{pmatrix} \xi_t \\ \varepsilon_t \end{pmatrix} \sim \text{ i.i.d. } N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix}\right).$$

Comparing Tables 4.4 and 4.5, it is apparent that separating consumption from dividends has a substantial effect on raising the premium, and on increasing return volatilities. This is not qualitatively entirely surprising, since there is now much greater dividend volatility — which makes equities more volatile — and this in turn affects an agent's $V()$ utility much more significantly. Although stock returns

Table 4.5 The equity premium in a model with loss aversion and distinct consumption and dividend processes (all return measures in percent).

		$b_0 = 0.7$ $k = 3$	$b_0 = 2$ $k = 3$	$b_0 = 100$ $k = 3$	$b_0 = 0.7$ $k = 20$	$b_0 = 2$ $k = 10$	$b_0 = 100$ $k = 8$
$\log r^f$	0.58	3.79	3.79	3.79	3.79	3.79	3.79
$E \log r^P$	6.03	1.3	2.62	3.68	5.17	5.02	5.88
$\sigma \log r^P$	20.02	17.39	20.87	20.47	25.85	23.84	24.04
Sharpe ratio	0.3	0.07	0.13	0.18	0.20	0.21	0.24
$\rho\left(\frac{C_{t+1}}{C_t}, r_{t+1}^P\right)$	0.10	0.15	0.15	0.15	0.15	0.15	0.15
Average loss aversion		2.25	2.25	2.25	5.8	3.5	3.2

Notes: $g_D = g_C = 1.84$, $\sigma_C = 3.79$, $\sigma_D = 0.12$, $w = 0.15$, $\eta = 0.9$, $\gamma = 1$, $\beta = 0.98$, $\lambda = 2.25$.
Source: This table is drawn from Table III in Barberis et al. (2001).

are now less highly correlated with consumption (Table 4.5), this does not matter since the investor is concerned about stock market volatility per se. This latter fact more than compensates for the correlation diminution. By way of contrast, this feature is relatively unproductive in habit style models as the increased volatility of returns is offset by its reduced correlation with consumption; the net consequence is very little change in the premium. An increase in the parameter k, by increasing loss aversion, has a similarly salutary effect on the premium as in Table 4.4.

Loss aversion/narrow framing is an appealing idea, and Barberis et al. (2001) analyze its equilibrium asset pricing implications in a careful and thorough way. There is, however, a sense in which their study is premature. In particular, we as yet lack choice theoretic underpinnings and the aggregation properties are as yet unconfirmed (strict aggregation will not hold).

The difficulty that several model classes have collectively 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.

5

Non-Risk Based Explanations of the Equity Premium

In this section we review the literature that takes as given the findings in Mehra and Prescott (1985) and tries to account for the equity premium by *factors other than aggregate risk.* Much of this literature re-examines the appropriateness of the abstractions and assumptions made in our original paper. In particular, borrowing constraints, the appropriateness of using T-bills as a proxy for the intertemporal marginal rate of substitution of consumption, the abstraction from taxes, diversification and intermediation costs, regulations and agent heterogeneity. We consider each in turn and examine the impact on the equity premium.

5.1 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.[1] Some recent attempts to resolve the puzzle incorporating both borrowing constraints *and* con-

---

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

sumer heterogeneity appear promising. One approach, which departs from the representative agent model, has been proposed in Constantinides et al. (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. 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:

$$\max_{\left\{z_{t,i}^e, z_{t,i}^b\right\}} E\left(\sum_{i=0}^{2} \beta^i U(C_{t,i})\right) \tag{5.1}$$

s.t.

$$c_{t,0} + q_t^e z_{t,0}^e + q_t^b z_{t,0}^b \leq w^0 \tag{5.2}$$

$$c_{t,1} + q_{t+1}^e z_{t,1}^e + q_{t+1}^b z_{t,1}^b \leq \left(q_{t+1}^e + d_{t+1}\right) z_{t,0}^e + \left(q_{t+1}^b + b\right) z_{t,0}^b + w_{t+1}^1 \tag{5.3}$$

$$c_{t,2} \leq \left(q_{t+2}^e + d_{t+2}\right) z_{t,1}^e + \left(q_{t+2}^b + b\right) z_{t,1}^b$$

$c_{t,j}$ 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_t^b and q_t^e 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: $z^e_{t,0} = 0$ and $z^b_{t,0} = 0$.

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.[2] Second, consumers of one generation are prohibited from trading bonds and equity with consumers of an unborn generation. They build on the observation that 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.[3]

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.[4]

---

[2] Being homogeneous within their generation, consumers have no incentive to trade claims with consumers of their own generation.

[3] See Mankiw (1986) and Constantinides and Duffie (1996).

[4] This is precisely the reason 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 high, i.e., when marginal utility is low, hence their price is (relatively) low and their rate of return is high.

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 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.

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.

5.2   The Choice of a Riskless Asset

In the two decades since "The Equity Premium: A Puzzle" was published, the way we think about economic phenomena has undergone a fundamental change, influenced in large measure by the development of Real Business Cycle theory by Kydland and Prescott (1982). One finding of the real business cycle literature is that the real after-tax return on capital, that is the after tax capital income divided by the cost of reproducing the tangible capital stock averages 4.5 percent with modest variation over time. A key question is: what is the *empirical counterpart* to the household's real interest rate? Mehra and Prescott (1985) used the highly liquid T-bill rate, corrected for expected inflation as a proxy for the return on the riskless asset that is used by agents to smooth consumption. The appropriateness of this assumption was questioned by McGrattan and Prescott (2003) and more recently by Mehra and Prescott (2008a). The remainder of this section closely follows the discussion in these papers.

An assumption implicit in Mehra and Prescott (1985) is that agents use *both* equity and the riskless asset to *intertemporally* smooth *consumption*. This is a direct consequence of the first-order condition (below) for the representative household in their model, which saves by optimally allocating resources between equity and riskless debt.

$$0 = E_t \left[\frac{U_c(c_{t+s})}{U_c(c_t)} (r^e_{t,t+s} - r^d_{t,t+s}) \right]. \tag{5.4}$$

Equation (5.4) is the standard asset pricing equation in macroeconomics and finance. $U_c(c_{t+s})$ is the marginal utility of consumption at time $t + s$, $r^e_{t,t+s}$, and $r^d_{t,t+s}$ are, respectively, the return on equity and the riskless asset over the period t, $t + s$, and E_t is the expectation conditional on the agent's information set at time t.

Is it reasonable to assume that the return on T-bills is equal to the real interest rate that households' use to save for retirement and smooth consumption? Do households actually hold T-bills to finance their retirement? If this were empirically true, it would be reasonable to equate their expected marginal rate of substitution of consumption to the rate of return on T-bills.

This question *cannot be answered in the abstract*, without reference to the asset holdings of households. A natural next step then is to examine the assets held by households. Table 5.1 details these holdings for American households. The four big asset-holding categories of households are tangible assets, pension and life insurance holdings, equity (both corporate and non corporate), and debt assets.

In the year 2000, privately held government debt was only 0.30 GDP, a third of which was held by foreigners. The amount of interest bearing government debt with maturity less than a year was only 0.085 GDP, which is a small fraction of the total household net worth. Virtually no T-bills are directly owned by households.[5] Approximately one-third of the outstanding T-bills are held by foreigners, in particular foreign central banks, and two-thirds by American financial institutions.

Although there are large amounts of debt assets held, most of these are in the form of pension fund and life insurance reserves. Some are

Table 5.1 Household assets and liabilities (average of 2000 and 2005).

Assets (GDP)		Liabilities (GDP)	
Tangible household assets	1.65	Liabilities	0.7
Corporate equity	0.85	Net worth	4.15
Non corporate equity	0.5		
Pension and life insurance reserves	1.0		
Debt assets	0.85		
Total	4.85	Total	4.85

---

[5] Table B-89, Economic Report of the President (2005).

in the form of demand deposits for which free services are provided. Most of the government debt is held indirectly, though some is in form of savings bonds that people gift to their grandchildren.

Thus, much of intertemporal savings done in the form of debt assets, is in assets such as annuities and mortgage debt held in retirement accounts and as pension fund reserves. *Other assets, not T-bills are held to finance consumption when old.*

A natural question then is who holds T-bills and do the holders use them to intertemporally smooth consumption? A large fraction of the stock of short-term US T-bills are held by foreign central banks and are used to smooth exchange rates. These central banks *are not* using these assets to equate the marginal utility of consumption today to the expected discounted marginal utility of consumption tomorrow.

US households do hold a sizable quantity of close substitutes for T-bills. These including checking accounts, money market accounts, and bank CDs with maturity of a year or less. They are of the order of magnitude of 0.50 times GDP. However, these short-term liquid assets with low expected return are held, in most part, for reasons *other than saving for retirement.* These include:

5.2.1   Liquidity

A part of these assets are held for liquidity purposes and as precautionary balances. The latter serve as a substitute for insurance against idiosyncratic risk, which is both large and very costly to insure against because of the associated moral hazard.

When this is the case, in the household's maximization problem there is a Lagrange multiplier that is not zero and as a result the expected returns are less than that predicted by a theory that abstracts from the holding of liquid assets as a substitute for insurance against idiosyncratic risk.

One of the first attempts to quantify this liquidity premium was the work of Bansal and Coleman (1996). 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 percent 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 returns, this may offer an explanation for the observed equity premium.

5.2.2 Transaction Balances

There are large transaction costs associated with moving into and out of high yielding assets. A large component is record keeping. With assets whose price varies, there are capital gains or loses with *every sale* of these assets. Every gain or loss must be reported on the annual tax form. This is why money market accounts hold debt securities until they mature. Hence there are no capital gains or loses associated when an individual buys or sells shares in his money market account. All that the money market mutual fund need report to its shareholders is an annual statement of the interest they received on the 1099 form. This is a major reason that sizable quantities of currency and non-interest bearing demand deposits are held. We note that Ml is about 15 percent of GDP.

Brokerage charges were (and perhaps continue to be for some) another significant cost associated with getting into and out of higher yielding securities such as stocks and bonds. It is true that discount brokers now exist, but gaining the prerequisite knowledge is costly in terms of time. If the holding period is short, even for a risk neutral investor it makes no sense for the investor to buy and later sell stocks unless the expected return on stocks is huge.

Based on the above insights, both McGrattan and Prescott (2003) and Mehra (2007), Mehra and Prescott (2007, 2008b) conclude that T-bills and short-term debt are not a reasonable empirical counterpart of the household interest rate in Equation (5.4). Hence, it would be inappropriate to equate the return on these assets to the marginal rate of substitution for an important group of agents.

McGrattan and Prescott (2003), who are concerned with differences in the return on debt and equity held to finance retirement, use long term high grade municipal bonds as their measure of the long term household interest rate during the Gold Standard period 1880–1934. They note that as these bonds were tax exempt and were held directly by individuals, no further adjustment for taxes and intermediation costs is required. They not adjust for inflationary expectations arguing that prices were expected to be stable during this period. For the post World War II and Korean War period 1960–2002, they use "the highest grade corporate bonds which were held primarily in tax-deferred pension funds." To account for inflationary expectations they subtract the inflation rate for the previous 10 years. Since there were government regulations during the inter war years that impact upon bond returns (see below), they used NIPA capital returns for these years. Their results are plotted in Figure 5.1. The returns for the entire period are seen to be remarkably close to 4 percent, considerably higher than the 0.8 percent return for T-bills reported by Mehra and Prescott (1985).

Mehra and Prescott (2007, 2008a,b) argue that an inflation indexed default free bond portfolio that has a similar duration to a well-diversified equity portfolio would be a reasonable proxy for a risk-free
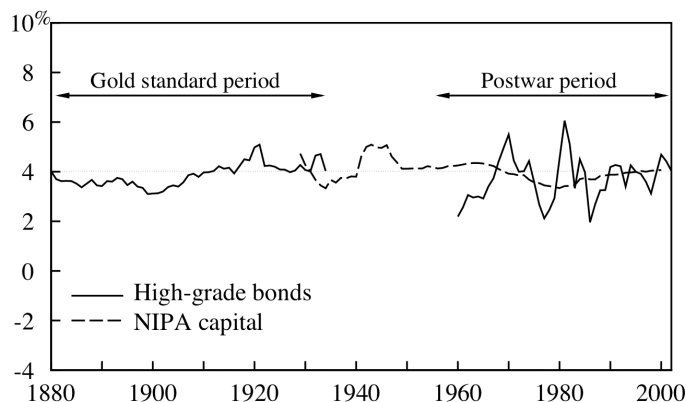

Fig. 5.1
Source: McGrattan and Prescott (2003).

asset used for consumption smoothing. For most of the 20th century, equity has had an implied duration of about 25 years so a portfolio of TIPS of a similar duration would be a reasonable proxy, but TIPS have only relatively recently (1997) been introduced in the US capital markets, and it is difficult to get accurate estimates of the mean return on this asset class. The average return for the 1997–2005 period was 3.7 percent. As an alternative (though imperfect) proxy, McGrattan and Prescott suggest the use of returns on indexed mortgages guaranteed by Ginnie Mae or issued by Fannie Mae. Using these indexed default free securities as a benchmark reduces the magnitude of the equity premium to something in the range of 2%–4%, depending on the benchmark used.

Recent work by Krishnamurthy and Vissing-Jorgensen (2007) provides evidence for a convenience yield on *all* treasury securities (not limited to T-bills). They document that the Debt/GDP ratio in the United States is negatively correlated with the spread between corporate bond yields and treasury bond yields and claim that the result holds even when controlling for the default risk of corporate bonds. They argue that the corporate bond spread reflects a convenience yield that investors attribute to Treasury debt. This yield, which varies with the Debt/GDP ratio, is in range of 0.2–1.6 percent.

This finding, if it holds up, provides additional evidence that returns on even long dated treasury securities probably underestimate the marginal rate of intertemporal substitution of consumption and further strengthen the arguments in McGrattan and Prescott (2003) and Mehra and Prescott (2007, 2008a,b).

5.3   The Effect of Government Regulations and Rules

McGrattan and Prescott (2003) argue that the estimated average return on debt assets in United States (including T-bills) over the period 1926–2000 is biased downwards because of various regulations (specifically W and X) that helped the treasury keep nominal rates below 2.5 percent during the 1941–1954 period. Table 5.2 shows that the return on debt securities during the 1941–1954 period was considerably lower than their long-term average value. This serves as a reminder

Table 5.2 US Inflation adjusted average return on debt.

Period	1926–2000 (%)	1926–1940 + 1955–2000 (%)	1941–1954 (%)
US treasury bills	0.8	1.8	−3.6
Intermediate-term government bonds	2.4	3.6	−2.7
Long-term government bonds	2.7	3.8	−1.9
Long-term corporate bonds	3.0	4.1	−1.9

Source: Ibbotson Associates (2006).

that governments can pursue regulatory policies that result in negative interest rates over an extended period of time. Clearly these rates have little to do with an agent's marginal rate of substitution that would be inferred were there no regulations. Such regulatory periods should be excluded in estimating the long-term average rates on debt securities.

The third column in Table 5.2 shows how the conventionally used numbers (in column two) change when the 1941–1954 period is excluded. The estimated average rates increase by about 1 percent for all asset classes.

Mehra and Prescott (2007, 2008a,b) point out that in the case of T-bills a further adjustment needs to be made to the returns in the 1930s. During that period, in some states, T-bills were exempt from personal property taxes while cash was not. This created an excess demand for the T-bills and they sold at a premium. Again, these rates of return have little to do with the marginal rate of substitution of consumption over time. The effect of these adjustments is to further reduce the magnitude of the equity premium relative to T-bills.

To summarize: using the return on a risk-free asset that is used for saving as a proxy for the intertemporal marginal rate of substitution of consumption (instead of a T-bill return) can significantly reduce the equity premium. Adjusting debt returns for government regulations further reduces the premium by 1 percent *irrespective* of the debt asset used as a benchmark.

5.4   Taxes

McGrattan and Prescott (2003) point out that although taxes on equity income were low until the mid 30s they rose dramatically during the war years. They argue that equity returns should be adjusted to reflect
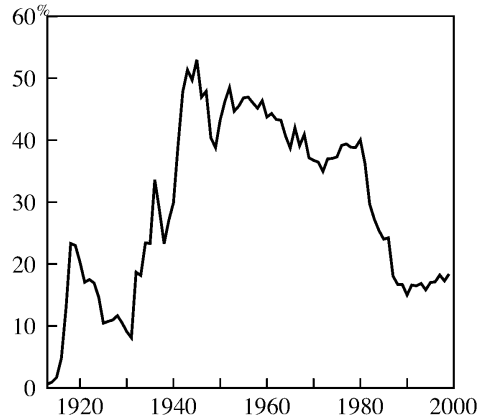
Fig. 5.2 The marginal tax rates on equity income distributions.
Source: McGrattan and Prescott (2003).

this, specially because of regulations that discouraged insurance companies and pension funds from holding equity in tax deferred accounts. They use IRS data to compute the marginal tax rates on equity income distributions. Figure 5.2 is a plot of the marginal tax rates on equity. These rates peaked at about 50 percent.

Consider the case where the tax on corporate distributions is increased from 0 percent to 50 percent and the additional revenue collected is lump summed back to households. If all investments are financed by retained earnings, as is nearly always the case, the value of the stock will be twice as high when the distribution tax rate is zero as when this tax rate is 50 percent. A consequence of this is that the return on equity computed in the standard way will be twice as high when the tax on distribution is 50 percent. This is important because there have been large changes in the tax rate on corporate distributions to owners. Figure 5.2 plots the distribution tax rate for the United States.

McGrattan and Prescott use this to compute the after tax equity returns to households after making a number of other adjustments to account for inflation, diversification costs and the fact that some equity was not taxed or was tax deferred. Figure 5.3 is a plot of their estimate of the after tax return on equity.
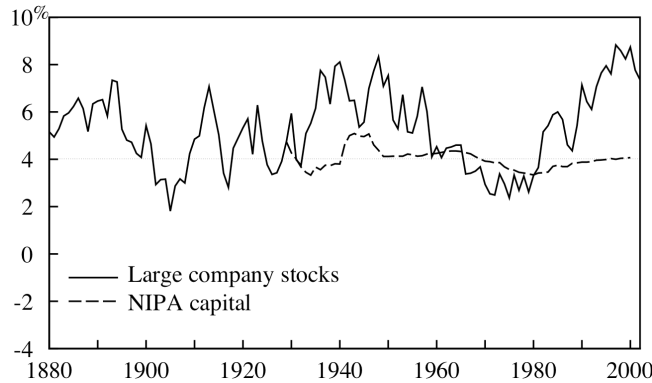
Fig. 5.3 Estimation of the after tax return on equity.
Source: McGrattan and Prescott (2003).

An interesting question that arises from their analysis is, why was the return on corporate equity so high in the 1960–2000 period? McGrattan and Prescott (2003, 2005) 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, if tax revenues are returned as a 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 be held tax free in 1962. In 2000, both could 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, 2005) in determining whether the stock market was overvalued or undervalued vis-à-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 US 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

$$V = (1 - \tau_d)K_T' + (1 - \tau_d)\left(1 - \tau_c\right)K_I' \tag{5.5}$$

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 at reproduction cost, and
K_I' is the end-of-period intangible corporate capital stock at reproduction cost.

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 the 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.[6]

The tricky part of the calculation is in constructing a measure of intangible capital. These investments reduce current accounting profits and the returns on the existing stock of intangible capital but increase future accounting profits. The formula for steady-state, pre-tax accounting profits is

$$\pi = \frac{i}{1 - \tau_c} \, K_T + i \, K_I - g \, K_I, \qquad (5.6)$$

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 (5.6) are reported in the system of national accounts with the exception of i and K_I.

McGrattan and Prescott (2005) estimate i using national income data. Their estimate of i is the after-tax real return on capital in the non corporate sector, which has as much capital as 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 US corporations grew in the period. An increase in the size of the corporate intangible capital stock was also a contributing factor.

McGrattan and Prescott (2005) 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

---

[6] In fact, formula (5.5) must be adjusted if economic depreciation and accounting depreciation are not equal and if there is an investment tax credit. See McGrattan and Prescott (2005).

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 et al. (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.

5.5   Agent Heterogeneity and Intermediation Costs

A limitation of the *homogenous household* construct is that it precludes the modeling of borrowing and lending amongst agents. In equilibrium, the shadow price of consumption at date $t + 1$ in terms of consumption at date t is such that the amount of borrowing and lending is zero. Homogenous household models are thus incapable of matching the *quantities* of assets held and *intermediated*.

To address this issue, Mehra and Prescott (2008b) construct a model economy that incorporates agent heterogeneity in the form of differences in the strength of the bequest motive. In light of their earlier finding (1985) that the premium for bearing non-diversifiable aggregate risk is small, their analysis abstracts from *aggregate risk*. The only uncertainty that agents face is idiosyncratic risk about the duration of their lifetime after retirement. Agents enter the retirement phase of their life cycle with wealth accumulated during their working years. They have identical preferences for consumption; however, they differ with respect to their intensity for bequests. In equilibrium, those with a strong bequest motive accumulate equity assets and, when retired, live

off the income of these assets. The households with no bequest motive buy annuities during their working years and consume the annuity benefits over their retirement years.

The incorporation of agent heterogeneity allows them to capture a key empirical fact — that there is a large amount of borrowing and lending between households, in particular, between older households. This borrowing is done either directly, by issuing mortgages to finance owner occupied housing or indirectly, by owning partially debt financed rental properties through direct or limited partnerships or REITS. They abstract from the small amount of borrowing and lending directly between households and assume that all of it is intermediated through financial institutions such as banks and pension funds. For the United States, in 2005 the amount intermediated was approximately 1.6 times the GDP.

The intermediation technology is constant returns to scale with intermediation costs being proportional to the amount intermediated. To calibrate the constant of proportionality, they use Flow of Funds statistics and data from National Income and Product Accounts. The calibrated value of this parameter equals the net interest income of financial intermediaries, divided by the quantity of intermediated debt and is a little over 2 percent.

In the absence of aggregate uncertainty, the return on equity and the borrowing rate are identical, since the agents who borrow are also marginal in equity markets. In their framework, government debt is not intermediated and thus its return is equal to the lending rate. The equity premium relative to government debt is the intermediation spread. The divergence between borrowing and lending rates gives rise to a 2 percent equity premium even in a world *without aggregate uncertainty.*

6

Concluding Comments

In this review, we have provided a glimpse of the vast literature on the equity premium puzzle. As a result of these research efforts, we have a deeper understanding of the role and importance of the abstractions that contribute to the puzzle. While no single explanation has fully resolved the anomaly, considerable progress has been made and the equity premium is a lesser puzzle today than it was 25 years ago.

Appendix

Expanding the set of technologies in a pure exchange, Arrow–Debreu economy to admit capital accumulation and production as in Brock (1979), Prescott and Mehra (1980) or Donaldson and Mehra (1984) does not increase the set of joint equilibrium processes on consumption and asset prices. Since the set of equilibria in a production company is a subset of those in an exchange economy, it follows immediately that if the equity premium cannot be accounted for in an exchange economy, modifying the technology to incorporate production will not alter this conclusion.[1]

To see this, let θ denote preferences, τ technologies, E the set of the exogenous processes on the aggregate consumption good, P the set of technologies with production opportunities, and $m(\theta, \tau)$ the set of equilibria for economy (θ, τ).

Theorem

$$\bigcup_{\tau \varepsilon E} m(\theta, \tau) \supset \bigcup_{\tau \varepsilon P} m(\theta, \tau)$$

---

[1] The discussion below is based on Mehra (1998).

Proof. For $\theta_0 \in \theta$, and $\tau_0 \in P$, let (a_0, c_0) be a joint equilibrium process on asset prices and consumption. A necessary condition for equilibrium is that the asset prices a_0 be consistent with c_0, the optimal consumption for the household with preferences θ_0. Thus, if (a_0, c_0) is an equilibrium then

$$a_0 = g(c_0, \theta),$$

where g is defined by the first-order necessary conditions for household maximization. This functional relation must hold for all equilibria, regardless of whether they are for a pure exchange or a production economy.

Let (a_0, c_0) be an equilibrium for some economy (θ_0, τ_0) with $\tau_0 \in P$. Consider the pure exchange economy with $\theta_1 = \theta_0$ and $\tau_1 = c_0$. Our contention is that (a_0, c_0) is a joint equilibrium process for asset prices and consumption for the pure exchange economy (θ_1, τ_1). For all pure exchange economies, the equilibrium consumption process is τ, so $c_1 = \tau_1 = c_0$, given more is preferred to less. If c_0 is the equilibrium process, the corresponding asset price must be $g(c_0, \theta_1)$. But $\theta_1 = \theta_0$ so $g(c_0, \theta_1) = g(c_0, \theta_0) = a_0$. Hence a_0 is the equilibrium for the pure exchange economy (θ_1, τ_1), proving the theorem. $\qquad\square$

Acknowledgments

References

Abel, A. B. (1988), 'Stock prices under time varying dividend risk: An exact solution in an infinite horizon general equilibrium model'. *Journal of Monetary Economics* **22**, 375–394.

Abel, A. B. (1990), 'Asset prices under habit formation and catching up with the joneses'. *A.E.R. Papers and Proceedings* **80**, 38–42.

Abel, A. B., N. Gregory Mankiw, L. H. Summers, and R. J. Zeckhauser (1989), 'Assessing dynamic efficiency: Theory and evidence'. *Review of Economic Studies* **56**, 1–20.

Ait-Sahalia, Y., J. A. Parker, and M. Yogo (2001), 'Luxury Goods and the Equity Premium'. Working Paper #8417, NBER.

Aiyagari, S. R. and M. Gertler (1991), 'Asset returns with transactions costs and uninsured individual risk'. *Journal of Monetary Economics* **27**, 311–331.

Altonji, J. G., F. Hayashi, and L. J. Kotlikoff (1992), 'Is the extended family altruistically linked?'. *American Economic Review* **82**, 1177–1198.

Alvarez, F. and U. Jermann (2000), 'Asset pricing when risk sharing is limited by default'. *Econometrica* **48**, 775–797.

Attanasio, O. P., J. Banks, and S. Tanner (2002), 'Asset holding and consumption volatility'. *Journal of Political Economy* **110**(4), 771–792.

Attanasio, O. P. and S. J. Davis (1997), 'Relative wage movements and the distribution of consumption'. *Journal of Political Economy* **104**, 1227–1262.

Auerbach, A. J. and L. J. Kotlikoff (1987), *Dynamic Fiscal Policy.* Cambridge University Press.

Azeredo, F. (2007), 'Essays on aggregate economics and finance'. Doctoral Dissertation. University of California, Santa Barbara.

Backus, D. K., B. R. Routledge, and S. E. Zin (2004), 'Exotic preferences for macroeconomists'. In: M. Gertler and K. Rogoff (eds.): *NBER Macroeconomics Annual 2004.* Cambridge, MA: MIT Press.

Bansal, R. and J. W. Coleman (1996), 'A monetary explanation of the equity premium, term premium and risk free rate puzzles'. *Journal of Political Economy* **104**, 1135–1171.

Bansal, R. and A. Yaron (2004), 'Risks for the long run: A potential resolution of asset pricing puzzles'. *Journal of Finance* **59**, 1481–1509.

Barberis, N. and M. Huang (2008), 'The loss aversion/narrow framing approach to the stock market pricing and participation puzzles'. In: R. Mehra (ed.): *Handbook of the Equity Risk Premium.* Elsevier, Amsterdam.

Barberis, N., M. Huang, and T. Santos (2001), 'Prospect theory and asset prices'. *The Quarterly Journal of Economics* **CXVI**(1), 1–53.

Barro, R. (2006), 'Rare disasters and asset markets in the twentieth century'. *Quarterly Journal of Economics* **121**, 823–866.

Basak, S. and D. Cuoco (1998), 'An equilibrium model with restricted stock market participation'. *The Review of Financial Studies* **11**, 309–341.

Becker, G. S. and R. J. Barro (1988), 'A reformulation of the economic theory of fertility'. *Quarterly Journal of Economics* **103**(1), 1–25.

Benartzi, S. and R. H. Thaler (1995), 'Myopic loss aversion and the equity premium puzzle'. *Quarterly Journal of Economics* **110**, 73–92.

Bewley, T. F. (1982), 'Thoughts on tests of the intertemporal asset pricing model'. Working Paper, Northwestern University.

Boldrin, M. L., L. J. Christiano, and J. D. M. Fisher (2001), 'Habit persistence, asset returns, and the business cycle'. *American Economic Review* **91**, 149–166.

Bonomo, M. and R. Garcia (1993), 'Disappointment aversion as a solution to the equity premium and the risk-free rate puzzles'. Working Paper, Université de Montréal.

Brav, A., G. M. Constantinides, and C. C. Geczy (2002), 'Asset pricing with heterogeneous consumers and limited participation: Empirical evidence'. *Journal of Political Economy* **110**, 793–824.

Breeden, D. (1979), 'An intertemporal asset pricing model with stochastic consumption and investment opportunities'. *Journal of Financial Economics* **7**, 265–296.

Brock, W. A. (1979), '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 and Trade*. New York: Academic Press.

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

Campbell, J. Y. (1999), 'Asset prices, consumption, and the business cycle'. In: J. B. Taylor and M. Woodford (eds.): *Handbook of Macroeconomics*, Vol. 1. Amsterdam: North–Holland, Chapt. 19, pp. 1231–1303.

Campbell, J. Y. (2001), 'Asset pricing at the millennium'. *Journal of Finance* **55**, 1515–1567.

Campbell, J. Y. and J. H. Cochrane (1999), 'By force of habit: A consumption – based explanation of aggregate stock market behavior'. *Journal of Political Economy* **107**, 205–251.

Cochrane, J. H. (1991), 'A simple test of consumption insurance'. *Journal of Political Economy* **99**, 957–976.

Cochrane, J. H. (1997), 'Where is the market going? uncertain facts and novel theories'. *Economic Perspectives* **21**, 3–37.

Cochrane, J. H. (2005), *Asset Pricing*. Princeton, NJ: Princeton University Press.

Cogley, T. (1999), 'Idiosyncratic risk and the equity premium: Evidence from the consumer expenditure survey'. Working Paper, Arizona State University.

Constantinides, G. M. (1990), 'Habit formation: A resolution of the equity premium puzzle'. *Journal of Political Economy* **98**, 519–543.

Constantinides, G. M. (2002), 'Rational asset prices'. *Journal of Finance* **57**, 1567–1591.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2002), 'Junior can't borrow: A new perspective on the equity premium puzzle'. *Quarterly Journal of Economics* **118**, 269–296.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2005), 'Junior must pay: Pricing the implicit put in privatizing social security'. *Annals of Finance* **1**, 1–34.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2007), 'Junior is rich: Bequests as consumption'. *Economic Theory* **32**, 125–155.

Constantinides, G. M. and D. Duffie (1996), 'Asset pricing with heterogeneous consumers'. *Journal of Political Economy* **104**, 219–240.

Cox, J. C., J. E. Ingersoll Jr., and S. A. Ross (1985), 'A theory of the term structure of interest rates'. *Econometrica* **53**, 385–407.

Daniel, K. and D. Marshall (1997), 'The equity premium puzzle and the risk-free rate puzzle at long horizons'. *Macroeconomic Dynamics* **l**, 452–484.

Danthine, J.-P. and J. B. Donaldson (2005), *Intermediate Financial Theory*. Upper Saddle River, NJ: Prentice Hall.

Danthine, J.-P., J. B. Donaldson, and R. Mehra (1992), 'The equity premium and the allocation of income risk'. *Journal of Economic Dynamics and Control* **16**, 509–532.

Davis, S. J. and P. Willen (2000), 'Using financial assets to hedge labor income risk: Estimating the benefits'. Working Paper, University of Chicago.

Dimson, E., P. Marsh, and M. Staunton (2002), *Triumph of the Optimists: 101 Years of Global Investment Returns*. Princeton, NJ: Princeton University Press.

Donaldson, J. B. and R. Mehra (1984), 'Comparative dynamics of an equilibrium intertemporal asset pricing model'. *Review of Economic Studies* **51**, 491–508.

Donaldson, J. B. and R. Mehra (2008), 'Risk based explanations of the equity premium'. In: R. Mehra (ed.): *Handbook of the Equity Risk Premium.* Elsevier, Amsterdam.

Duffie, D. (2001), *Dynamic Asset Pricing Theory.* Princeton, NJ: Princeton University Press, 3rd edition.

Economic Report of the President (2005), *Council of Economic Advisors.* United States Government Printing Office, Washington.

Epstein, L. G. and S. E. Zin (1991), 'Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis'. *Journal of Political Economy* **99**, 263–286.

Ferson, W. E. and G. M. Constantinides (1991), 'Habit persistence and durability in aggregate consumption'. *Journal of Financial Economics* **29**, 199–240.

Gabaix, X. and D. Laibson (2001), 'The 6D bias and the equity premium puzzle'. In: B. Bernanke and K. Rogoff (eds.): *NBER Macroeconomics Annual 2001.* Cambridge, MA: 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.

Gul, F. (1991), 'A theory of disappointment aversion'. *Econometrica* **59**(3), 667– 686.

Hansen, L. P. and R. Jagannathan (1991), 'Implications of security market data for models of dynamic economies'. *Journal of Political Economy* **99**, 225–262.

Hansen, L. P. and K. J. Singleton (1982), 'Generalized instrumental variables estimation of nonlinear rational expectations models'. *Econometrica* **50**, 1269–1288.

He, H. and D. M. Modest (1995), 'Market frictions and consumption-based asset pricing'. *Journal of Political Economy* **103**, 94–117.

Heaton, J. (1995), 'An empirical investigation of asset pricing with temporally dependent preference specifications'. *Econometrica* **66**, 681–717.

Heaton, J. and D. J. Lucas (1996), 'Evaluating the effects of incomplete markets on risk sharing and asset pricing'. *Journal of Political Economy* **104**, 443–487.

Heaton, J. and D. J. Lucas (1997), 'Market frictions, savings behavior and portfolio choice'. *Journal of Macroeconomic Dynamics* **1**, 76–101.

Heaton, J. C. and D. J. Lucas (2000), 'Portfolio choice and asset prices: The importance of entrepreneurial risk'. *Journal of Finance* **55**, 1163–1198.

Ibbotson Associates (2006), Stocks, Bonds, Bills and Inflation. *2005 Yearbook*. Chicago: Ibbotson Associates.

Jacobs, K. (1999), 'Incomplete markets and security prices: Do asset-pricing puzzles result from aggregation problems?'. *Journal of Finance* **54**, 123–163.

Jacobs, K. and K. Q. Wang (2004), 'Idiosyncratic consumption risk and the cross-section of asset returns'. *Journal of Finance* **59**, 2211–2252.

Kahneman, D. and A. Tervsky (1979), 'Prospect theory: An analysis of decision under risk'. *Econometrica* **47**(2), 263–292.

Kandel, S. and R. F. Stambaugh (1991), 'Asset returns and intertemporal preferences'. *Journal of Monetary Economics* **27**, 39–71.

Kocherlakota, N. R. (1996), 'The equity premium: It's still a puzzle'. *Journal of Economic Literature* **34**, 42–71.

Krebs, T. (2000), 'Consumption-based asset pricing with incomplete markets'. Working Paper, Brown University.

Krishnamurthy, A. and A. Vissing-Jorgensen (2007), 'The Demand for Treasury Debt'. Working Paper, Northwestern University.

Kydland, F. and E. C. Prescott (1982), 'Time to build and aggregate fluctuations'. *Econometrica* **50**, 1345–1371.

LeRoy, S. H. and J. Werner (2001), *Principles of Financial Economics*. New York, NY: Cambridge University Press.

Ljungqvist, L. and H. Uhlig (1999), 'On consumption bunching under campbell–cochrane habit formation'. SFB 649 Discussion Paper, Berlin.

Lucas, D. J. (1994), 'Asset pricing with undiversifiable risk and short sales constraints: Deepening the Equity premium puzzle'. *Journal of Monetary Economics* **34**, 325–341.

Lucas Jr, R. E. (1978), 'Asset prices in an exchange economy'. *Econometrica* **46**, 1429–1445.

Luttmer, E. G. J. (1996), 'Asset pricing in economies with frictions'. *Econometrica* **64**, 1439–1467.

Lynch, A. W. (1996), 'Decision frequency and synchronization across agents: Implications for aggregate consumption and equity returns'. *Journal of Finance* **51**, 1479–1497.

Mace, B. J. (1991), 'Full insurance in the presence of aggregate uncertainty'. *Journal of Political Economy* **99**, 928–956.

Mankiw, N. G. (1986), 'The equity premium and the concentration of aggregate shocks'. *Journal of Financial Economics* **17**, 211–219.

Mankiw, N. G. and S. P. Zeldes (1991), 'The consumption of stockholders and nonstockholders'. *Journal of Financial Economics* **29**, 97–112.

McGrattan, E. R. and E. C. Prescott (2000), 'Is the stock market overvalued?'. *Federal Reserve Bank of Minneapolis Quarterly Review.*

McGrattan, E. R. and E. C. Prescott (2003), 'Average debt and equity returns: Puzzling?'. *American Economic Review* **93**, 392–397.

McGrattan, E. R. and E. C. Prescott (2005), 'Taxes, regulations, and the value of U.S. and U.K. corporations'. *Review of Economic Studies* **92**, 767–796.

Mehra, R. (1988), 'On the existence and representation of equilibrium in an economy with growth and nonstationary consumption'. *International Economic Review* **29**, 131–135.

Mehra, R. (1998), 'On the volatility of stock prices: An exercise in quantitative theory'. *International Journal of Systems Science* **29**, 1203–1211.

Mehra, R. (2003), 'The equity premium: Why is it a puzzle'. *Financial Analysts Journal* **59**, 54–69.

Mehra, R. (2007), 'The equity premium in India'. In: K. Basu (ed.): *Oxford Companion to Economics in India.* Oxford University Press.

Mehra, R. (ed.) (2008), 'Handbook of the Equity Risk Premium'. Elsevier, Amsterdam.

Mehra, R. and E. C. Prescott (1985), 'The equity premium: A puzzle'. *Journal of Monetary Economics* **15**, 145–161.

Mehra, R. and E. C. Prescott (1988), 'The equity premium: A solution?'. *Journal of Monetary Economics* **22**, 133–136.

Mehra, R. and E. C. Prescott (2003), 'The equity premium in retrospect'. In: G. M. Constantinides, M. Harris, and R. Stulz (eds.): *Handbook of the Economics of Finance.* Amsterdam, Netherlands: North-Holland.

Mehra, R. and E. C. Prescott (2007), 'Intermediated quantities and returns'. Working Paper, UCSB.

Mehra, R. and E. C. Prescott (2008a), 'The equity premium: ABCs'. In: R. Mehra (ed.): *Handbook of the Equity Risk Premium.* Elsevier, Amsterdam.

Mehra, R. and E. C. Prescott (2008b), 'Non-Risk Based Explanations of the Equity Premium'. In: R. Mehra (ed.): *Handbook of the Equity Risk Premium.* Elsevier, Amsterdam.

Merton, R. C. (1971), 'Optimum consumption and portfolio rules in a continuous time model'. *Journal of Economic Theory* **3**, 373–413.

Mossin, J. (1966), 'Equilibrium in a capital asset market'. *Econometrica* **34**(4), 768–783.

Prescott, E. C. and R. Mehra (1980), 'Recursive competitive equilibrium: The case of homogeneous households'. *Econometrica* **48**, 1365–1379.

Rietz, T. A. (1988), 'The equity risk premium: A solution'. *Journal of Monetary Economics* **22**, 117–131.

Routledge, B. and S. Zin (2004), *Generalized Disappointment Aversion and Asset Prices.* Mimeo: Carnegie-Mellon University.

Rubinstein, M. (1976), 'The valuation of uncertain income streams and the pricing of options'. *Bell Journal of Economics* **7**, 407–425.

Shiller, R. J. (1989), 'Comovements in stock prices and comovements in dividends'. *Journal of Finance* **44**, 719–729.

Siegel, J. (2002), *Stocks for the Long Run.* New York: Irwin, 2nd edition.

Storesletten, K., C. I. Telmer, and A. Yaron (2004), 'Consumption and risk sharing over the life cycle'. *Journal of Monetary Economics* **51**(3), 609–633.

Storesletten, K., C. I. Telmer, and A. Yaron (2007), 'Asset pricing with idiosyncratic risk and overlapping generations'. *Review of Economic Dynamics.* Forthcoming.

Telmer, C. I. (1993), 'Asset – pricing puzzles and incomplete markets'. *Journal of Finance* **49**, 1803–1832.

Vissing-Jorgensen, A. (2002), 'Limited asset market participation and the elasticity of intertemporal substitution'. *Journal of Political Economy* **110**, 825–853.

Weil, P. (1989), 'The equity premium puzzle and the risk – free rate puzzle'. *Journal of Monetary Economics* **24**, 401–421.