Discussion Paper 102

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480-0291

August 1995

The Equity Premium: It's Still a Puzzle

Narayana R. Kocherlakota*

University of Iowa

ABSTRACT

The paper examines the literature that attempts to resolve the equity premium and riskfree rate puzzles. It demonstrates that the puzzles will confront any model of asset prices that relies on three crucial assumptions: preferences have a particular parametric form, asset markets are complete, and asset trade is frictionless. A survey of the literature that relaxes these assumptions reveals that there are now several plausible explanations of the seemingly low riskfree rate, but the large size of the equity premium remains a puzzle.

*I thank Rao Aiyagari, John Campbell, Dean Corbae, John Heaton, Mark Huggett, Beth Ingram, John Kennan, Deborah Lucas, Barbara McCutcheon, Gene Savin, Steve Williamson, Kei-Mu Yi, the referees and the editor for their input in preparing this manuscript. I thank Edward R. Allen III, Ravi Jagannathan, and Deborah Lucas for our many interesting and enlightening conversations on the subject of asset pricing over the years. Finally, I thank Lars Hansen for teaching me how to think about the issues addressed in this paper. None of the preceding should be viewed as being responsible for the opinions and errors contained within. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

I. INTRODUCTION

Over the last one hundred years, the average real return to stocks in the United States has been about 6% per year higher than that on Treasury bills. At the same time, the average real return on Treasury bills has been about 1% per year. In this paper, I discuss and assess various theoretical attempts to explain these two different empirical phenomena: the large "equity premium" and the low "riskfree rate". I show that while there are several plausible explanations for the low level of Treasury returns, the large equity premium is still largely a mystery to economists.

In order to understand why the sample means of the equity premium and the riskfree rate represent "puzzles", it is useful to review the basics of modern asset pricing theory. It is well-known that the real returns paid by different financial securities may differ considerably, even when averaged over long periods of time. Financial economists typically explain these differences in average returns by attributing them to differences in the degree to which a security's return covaries with the typical investor's consumption. If this covariance is high, selling off the security would greatly reduce the variance of the typical investor's consumption stream; in equilibrium, the investor must be deterred from reducing his risk in this fashion by the security's paying a high average return.

There is a crucial problem in making this qualitative explanation of cross-sectional differences in asset returns operational: what exactly is "the typical investor's consumption"? The famous Capital Asset Pricing Model represents one answer to this question: it assumes that the typical investor's consumption stream is perfectly correlated with the return to the stock market. This allows financial analysts to measure the risk of a financial security by its covariance with the return to the stock market (that is, a scaled version of its "beta"). More recently, Douglas Breeden (1979) and Robert Lucas (1978) described so-called "representative" agent models of asset returns in which per capita consumption is perfectly correlated with the consumption stream of the typical investor[1]. In this type of model, a

[1]See also Mark Rubinstein (1976) and Breeden and Robert Litzenberger (1978).

security's risk can be measured using the covariance of its return with per capita consumption.

The representative agent model of asset pricing is not as widely used as the CAPM in "real world" applications (such as project evaluation by firms). However, from an academic economist's perspective, the "representative agent" model of asset pricing is more important than the CAPM. Representative agent models that imbed the Lucas-Breeden paradigm for explaining asset return differentials are an integral part of modern macroeconomics and international economics. Thus, any empirical defects in the representative agent model of asset returns represent holes in our understanding of these important subfields.

In their seminal (1985) paper, Rajnish Mehra and Edward Prescott describe a particular empirical problem for the representative agent paradigm. As mentioned above, over the last century, the average annual real return to stocks has been about 7% per year, while the average annual real return to Treasury bills has only been about 1% per year. Mehra and Prescott (1985) show that the difference in the covariances of these returns with consumption growth is only large enough to explain the difference in the average returns if the typical investor is implausibly averse to risk. This is the *equity premium puzzle*: in a *quantitative* sense, stocks are not sufficiently riskier than Treasury bills to explain the spread in their returns.

Philippe Weil (1989) shows that the same data presents a second anomaly. According to standard models of individual preferences, when individuals want consumption to be smooth over states (they dislike risk), they also desire smoothness of consumption over time (they dislike growth). Given that the large equity premium implies that investors are highly risk averse, the standard models of preferences would in turn imply that they do not like growth very much. Yet, although Treasury bills offer only a low rate of return, individuals defer consumption (that is, save) at a sufficiently fast rate to generate average per capita consumption growth of around 2% per year. This is what Weil (1989) calls the *riskfree rate puzzle*: although individuals like consumption to be very smooth, and although the riskfree rate is very low, they still save enough that per capita consumption grows rapidly.

There is now a vast literature that seeks to resolve these two puzzles. I begin my review of this literature by showing that the puzzles are very robust: they are implied by only three assumptions about individual behavior and asset market structure. First, individuals have *preferences* associated with the "standard" utility function used in macroeconomics: they maximize the expected discounted value of a stream of utilities generated by a power utility function. Second, asset markets are *complete*: individuals can write insurance contracts against any possible contingency. Finally, asset trading is *costless*, so that taxes and brokerage fees are assumed to be insignificant. Any model that is to resolve the two puzzles must abandon at least one of these three assumptions.

My survey shows that relaxing the three assumptions has led to plausible explanations for the low value of the riskfree rate. Several alternative preference orderings are consistent with it; also, the existence of borrowing constraints push interest rates down in equilibrium. However, the literature only provides two rationalizations for the large equity premium: either investors are highly averse to consumption risk and/or they find trading stocks to be much more costly than trading bonds. Little auxiliary evidence exists to support either of these explanations, and so (I would say) the equity premium puzzle is still unresolved.

The Lucas-Breeden representative agent model is apparently inconsistent with the data in many other respects. Sanford Grossman and Robert Shiller (1981) (and many others) argue that asset prices vary too much to be explained by the variation in dividends or in per capita consumption. Lars Hansen and Kenneth Singleton (1982, 1983) point out that in forming her portfolio, the representative investor appears to be ignoring useful information available in lagged consumption growth and asset returns. Robert Hall (1988) argues that consumption does not respond sufficiently to changes in expected returns. John Cochrane and Hansen (1992) uncover several anomalies, including the so-called Default-Premium and Term Structure Puzzles.

This paper will look at none of these other puzzles. Instead, it focuses exclusively on the riskfree rate and equity premium puzzles. There are two rationalizations for limiting the discussion in this way. The first is simple: my conversations with other economists have convinced me that these puzzles are the most widely known and best understood of the variety of evidence usually arrayed against the representative agent models.

The second rationalization is perhaps more controversial: I claim that these two puzzles have more importance for macroeconomists. The riskfree rate puzzle indicates that we do not know why people save even when returns are low: thus, our models of aggregate savings behavior are omitting some crucial element. The equity premium puzzle demonstrates that we do not know why individuals are so averse to the highly procyclical risk associated with stock returns. As Andrew Atkeson and Christopher Phelan (1994) point out, without this knowledge we cannot hope to give a meaningful answer to R. Lucas' (1987) question about how costly individuals find business cycle fluctuations in consumption growth. Thus, the two puzzles are signs of large gaps in our understanding of the macroeconomy.

Given the lack of a compelling explanation for the large equity premium, the article will at times read like a litany of failure. I should say from the outset that this is misleading in many ways. We have learned much from the equity premium puzzle literature about the properties of asset pricing models, about methods of estimating and testing asset pricing models, and about methods of solving for the implications of asset pricing models. I believe that all of these contributions are significant.

For better or for worse, though, this article is not about these innovations. Instead, I focus on what might be called the bottom line of the work, "Do we know why the equity premium is so high? Do we know why the riskfree rate is so low?" It is in answering the first of these questions that the literature falls short.

The rest of the paper is structured as follows. The next section describes the two puzzles and lays out the fundamental modelling assumptions that generate them. Section III explores the potential for explaining the two puzzles by changing the preferences of the representative agent. Section IV looks at the implications of market frictions for asset returns. Section V concludes[2].

II. WHAT ARE THE PUZZLES?

1. Aspects of the Data

The equity premium and riskfree rate puzzles concern the co-movements of three variables: the real return to the S & P 500, the real return to short-term nominally riskfree bonds[3], and the growth rate of per capita real consumption (more precisely, nondurables and services). In this paper, I use the annual United States data from 1889-1978 originally studied by Mehra and Prescott (1985) (see the Appendix for details on the construction of the data)[4]. Plots of the data are depicted in Figures 1-3; all of the series appear stationary and ergodic (their statistical properties do not appear to be changing over time).

Table 1 contains some summary statistics for the three variables. There are three features of the table that give rise to the equity premium and riskfree rate puzzles. First, the average real rate of return on stocks is equal to 7% per year while the average real rate of return on bonds is equal to 1% per year. Second, by long term historical standards, per capita consumption growth is high: around 1.8% per year.

---

[2]In writing this paper, I have benefitted greatly from reading four other review articles by Andrew Abel (1991), S. Rao Aiyagari (1993), Cochrane and Hansen (1992), and John Heaton and Deborah Lucas (1995b). I recommend all of them highly.

[3]Ninety day Treasury bills from 1931-1978, Treasury certificates from 1920-31, and sixty day to ninety day Commercial Paper prior to 1920.

[4]In what follows, I only use the Mehra-Prescott data set. However, it is important to realize that the puzzles are *not* peculiar to these data. For example, Kocherlakota (1994) shows that adding ten more years of data to the Mehra-Prescott series does not eliminate the puzzles. Hansen and Singleton (1983) and Aiyagari (1993) find that similar phenomena characterize post-World War II monthly data in the United States. Amlan Roy (1994) documents the existence of the two puzzles in post-World War II quarterly data in Germany and Japan. It may not be too strong to say that the equity premium and riskfree rate puzzles appear to be a general feature of organized asset markets. (Jeremy Siegel (1992) points out that the equity premium was much lower in the 19th century. However, consumption data is not available for this period.)

Table 1

Summary Statistics
United States Annual Data, 1889-1978

Sample Means

R_t^s	0.070
R_t^b	0.010
C_t/C_{t-1}	0.018

Sample Variance-Covariance

	R_t^s	R_t^b	C_t/C_{t-1}
R_t^s	0.0274	0.00104	0.00219
R_t^b	0.00104	0.00308	-0.000193
C_t/C_{t-1}	0.00219	-0.000193	0.00127

In this table, C_t/C_{t-1} is real per capita consumption growth, R_t^s is the real return to stocks and R_t^b is the real return to Treasury bills.

Figure 1: Annual Real Per Capita Consumption Growth

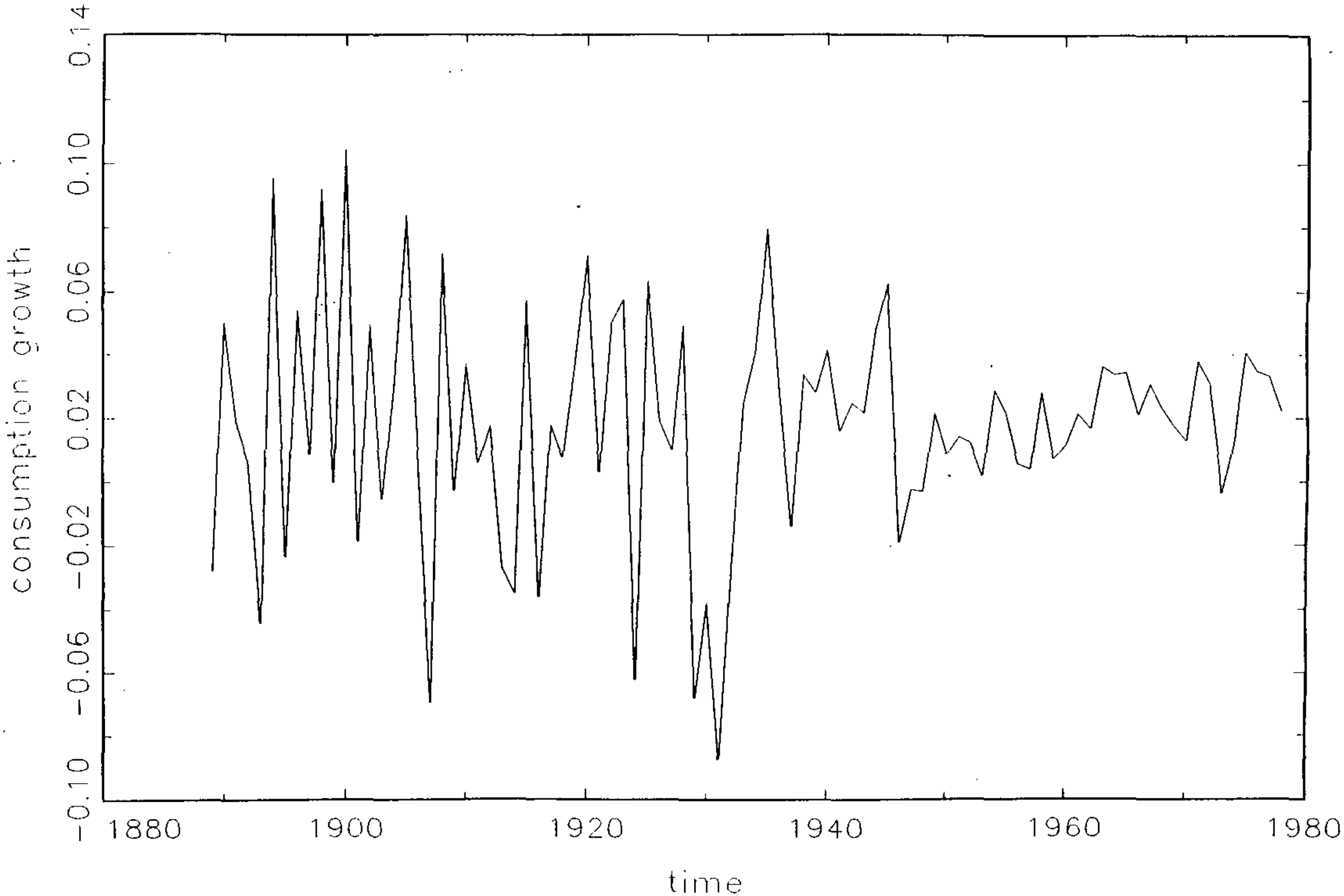


Figure 2: Annual Real Return to S & P 500

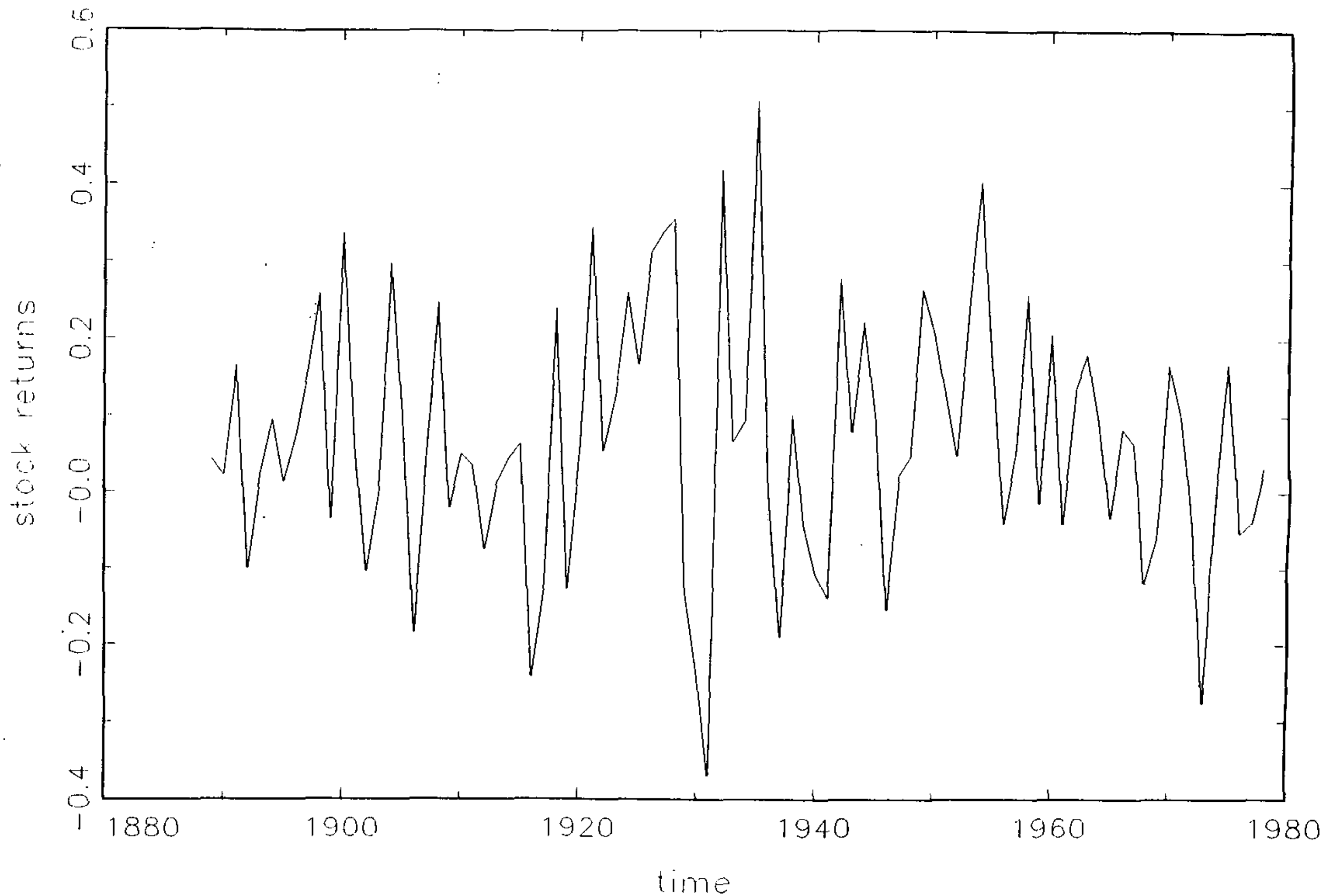
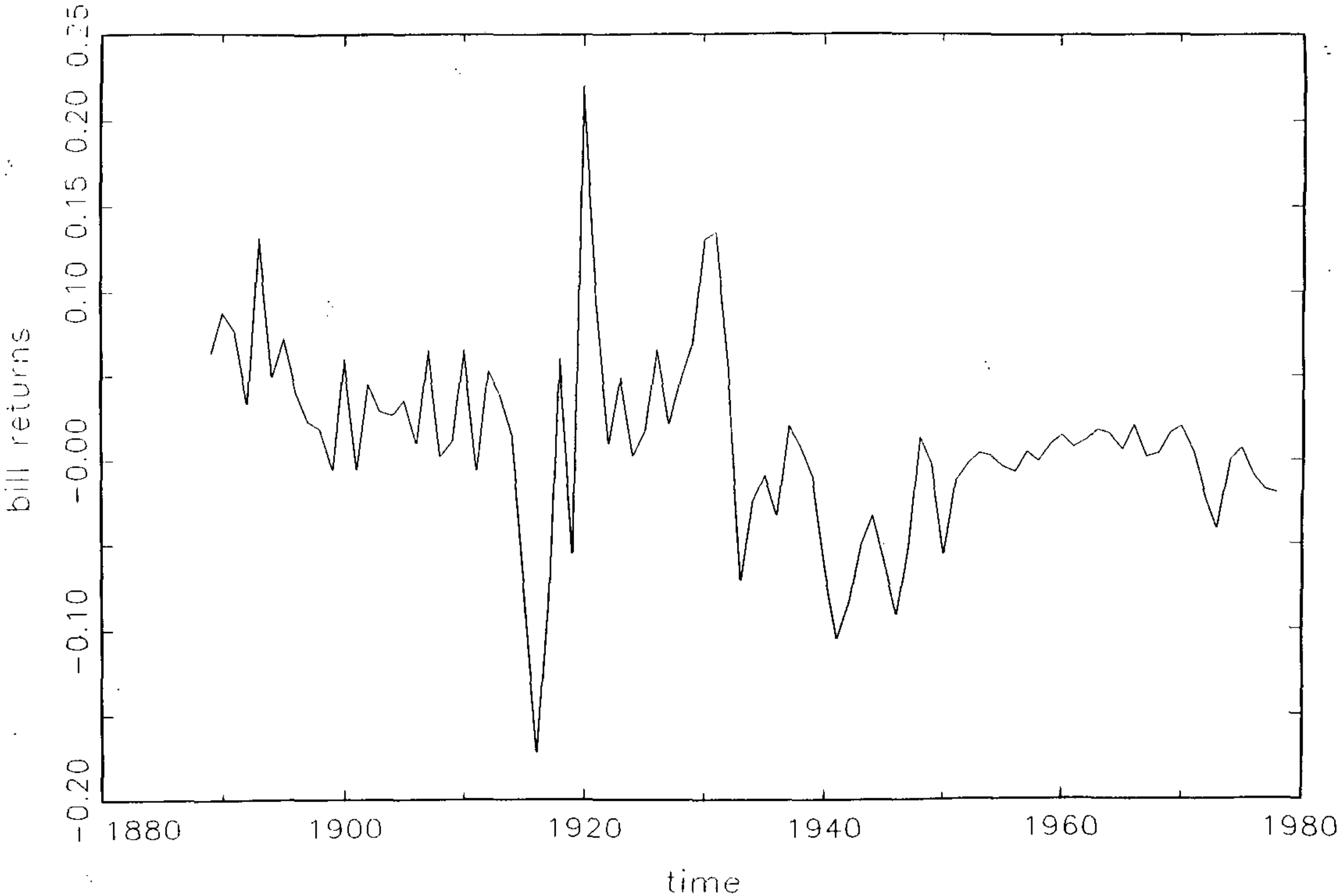


Figure 3: Annual Real Return to Nominally Riskfree Short Term Debt



Finally, the covariance of per capita consumption growth with stock returns is only slightly bigger than the covariance of per capita consumption growth with bond returns.

In some sense, there is a simple explanation for the higher average return of stocks: stock returns covary more with consumption growth than do Treasury bills. Investors see stocks as a poorer hedge against consumption risk, and so stocks must earn a higher average return. Similarly, while individuals must have saved a lot to generate the high consumption growth described in Table 1, there is an explanation of why they did so: as long as individuals discount the future at a rate lower than 1% per year, we should expect their consumptions to grow. The data does not contradict the *qualitative* predictions of economic theory; rather, as we shall see, it is inconsistent with the *quantitative* implications of a particular economic model.

2. The Original Statement of the Puzzles

In their 1985 paper, Mehra and Prescott construct a simple model that makes well-defined quantitative predictions for the expected values of the real returns to the S & P 500 and the three month Treasury bill rate. There are three critical assumptions underlying their model. The first is that in period t, all individuals have identical preferences over future random consumption streams represented by the utility function:

$$(1) \qquad E_t \Sigma_{s=0}^{\infty} \beta^s (c_{t+s})^{1-\alpha}/(1-\alpha), \ \alpha \geq 0$$

where $\{c_t\}_{t=1}^{\infty}$ is a random consumption stream. In this expression, and throughout the remainder of the paper, E_t represents an expectation conditional on information available to the individual in period t. Thus, individuals seek to maximize the expectation of a discounted flow of utility over time.

In the formula (1), the parameter β is a discount factor that households apply to the utility derived from future consumption; increasing β leads investors to save more. In contrast, increasing α has two seemingly distinct implications for individual attitudes towards consumption profiles. When α is large, individuals want consumption in different *states* to be highly similar: they dislike risk. But individuals also want consumption in different *dates* to be similar: they dislike growth in their consumption profiles.

The second key assumption of the Mehra-Prescott model is that people can trade stocks and bonds in a *frictionless* market. What this means is that an individual can costlessly (that is, without significant taxes or brokerage fees) buy and sell any amount of the two financial assets. In equilibrium, an individual must not be able to gain utility at the margin by selling bonds and then investing the proceeds in stocks; the individual should not be able to gain utility by buying or selling bonds. Hence, given the costlessness of performing these transactions, an individual's consumption profile must satisfy the following two first order conditions:

(2a) $$E_t\{(c_{t+1}/c_t)^{-\alpha}(R^s_{t+1} - R^b_{t+1})\} = 0$$

(2b) $$\beta E_t\{(c_{t+1}/c_t)^{-\alpha}R^b_{t+1}\} = 1$$

In these conditions, R^s_t is the gross return to stocks from period (t-1) to period t, while R^b_t is the gross return to bonds from period (t-1) to period t.

These first order conditions impose statistical restrictions on the comovement between any person's pattern of consumption and asset returns. The third critical feature of Mehra and Prescott's analysis, though, is that they assume the existence of a "representative" agent. According to this assumption, the above conditions (2a) and (2b) are satisfied not just for each individual's consumption, but also for *per capita* consumption.

There is some confusion about what kinds of assumptions underlie this substitution of per capita for individual consumption. It is clear that if all individuals are identical in preferences and their ownership of production opportunities, then in equilibrium, all individuals will in fact consume the same amount; thus, under these conditions, substituting per capita consumption into conditions (2a) and (2b) is justified. Many economists distrust representative agent models because they believe this degree of homogeneity is unrealistic.

However, while this degree of homogeneity among individuals is *sufficient* to guarantee the existence of a representative agent, it is by no means *necessary*. In particular, George Constantinides (1982) shows that even if individuals are heterogeneous in preferences and levels of wealth, it may be possible to find some utility function for the representative individual (with coefficient of relative risk aversion no larger than the most risk averse individual and no smaller than the least risk averse individual) which satisfies the first order conditions (2a) and (2b). The key is that asset markets must be *complete*: individuals in the United States must have a sufficiently large set of assets available for trade that they can diversify any idiosyncratic risk in consumption. When markets are complete, we can construct a "representative" agent because after trading in complete markets, individuals become *marginally* homogeneous even though they are initially heterogeneous.

We can summarize this discussion as follows: Mehra and Prescott assume that asset markets are frictionless, that asset markets are complete, and that the resultant representative individual has preferences of the form given by (1). However, Mehra and Prescott also make three other, more technical, assumptions. First, they assume that per capita consumption growth follows a two state Markov chain constructed in such a way that the population mean, variance, and autocorrelation of consumption growth are equivalent to their corresponding sample means in United States data. They also assume that in period t, the only variables that individuals know are the realizations of current and past consumption growth. Finally, they assume that the growth rate of the total dividends paid by the stocks included in the S & P 500 is perfectly correlated with the growth rate of per capita consumption, and that the real return to the (nominally riskfree) Treasury bill is perfectly correlated with the return to a bond that is riskfree in *real* terms.

These statistical assumptions allow Mehra and Prescott to use the first order conditions (2a) and (2b) to obtain an analytical formula that expresses the population mean of the real return to the S & P 500 and the population mean real return to the three month Treasury bill in terms of the two preference parameters β and α. Using evidence from microeconometric data and introspection, they restrict β to lie between 0 and 1, and α to lie between 0 and 10. Their main finding is that for any value of the preference parameters such that the expected real return to the Treasury bill is less than 4%, the difference in the two expected real returns (the "equity premium") is less than 0.35%. This stands in contrast with the facts described earlier that the average real return to Treasury bills in the United States data is 1% while the average real return to stocks is nearly 6% higher.

Mehra and Prescott conclude from their analysis that their model of asset returns is inconsistent with United States data on consumption and asset returns. Since the model was constructed by making several different assumptions, presumably the fit of the model to the data could be improved by changing any of them. In the next section, I re-state the equity premium puzzle so as to show why the three final assumptions about the statistical behavior of consumption and asset returns are relatively unimportant.

3. A More Robust Restatement of the Puzzles

As we saw above, the crux of the Mehra-Prescott model is that the first order conditions (2a) and (2b) must be satisfied with per capita consumption growth substituted in for individual consumption growth. Using the Law of Iterated Expectations, we can replace the conditional expectation in (2a) and (2b) with an unconditional expectation so that:

(2a′) $$E\{(C_{t+1}/C_t)^{-\alpha}(R^s_{t+1} - R^b_{t+1})\} = 0$$

(2b′) $$\beta E\{(C_{t+1}/C_t)^{-\alpha}R^b_{t+1}\} = 1$$

(The variable C_t stands for per capita consumption.) We can estimate the expectations or population means on the left hand side of (2a′) and (2b′) using the sample means of:

$$e^s_{t+1} = \{(C_{t+1}/C_t)^{-\alpha}(R^s_{t+1} - R^b_{t+1})\}$$

$$e^b_{t+1} = \beta\{(C_{t+1}/C_t)^{-\alpha}R^b_{t+1}\}$$

given a sufficiently long time series of data on C_{t+1} and the asset returns[5].

Table 2 reports the sample mean of e^s_t for various values of α; for all values of α less than or equal to 8.5, the sample mean of e^s_t is significantly positive. We can conclude that the population mean of e^s_t is in fact positive for any value of α less than or equal to 8.5; hence, for any such α, the representative agent can gain at the margin by borrowing at the Treasury bill rate and investing in stocks[6]. This is the equity premium puzzle. Intuitively, while the covariance between stock returns and per capita consumption growth is positive (see Table 1), it is not sufficiently large to deter the representative investor with a coefficient of relative risk aversion less than 8.5 from wanting to borrow

[5]Throughout this paper, I will be evaluating various representative agent models by using versions of (2a′) and (2b′) as opposed to (2a) and (2b). While (2a,b) certainly imply (2a′,b′), there are a host of other implications of (2a,b): the law of iterated expectations "averages" over all of the implications of (2a) and (2b) to arrive at (2a′) and (2b′). Focusing on (2a′) and (2b′) is in keeping with my announced goal in the introduction of just looking at the equity premium and riskfree rate puzzles.

Hansen and Ravi Jagannathan (1991) and Cochrane and Hansen (1992) reinterpret the equity premium puzzle and the riskfree rate puzzles using the variance bounds derived in the former paper. Note that, as Cochrane and Hansen (1992) emphasize, the variance bound is an even weaker implications of (2a) than (2a′) is (in the sense that (2a′) implies the variance bound while the converse is not true).

[6]Breeden, Michael Gibbons and Litzenberger (1989) and Grossman, Angelo Melino and Shiller (1987) point out that the first order conditions (2a) and (2b) are based on the assumption that all consumption within a given year is perfectly substitutable; if this assumption is false, there is the possibility of time aggregation bias. However, Hansen, Heaton and Amir Yaron (1994) show that as long as the geometric average of consumption growth rates is a good approximation to the arithmetic average of consumption growth rates, the presence of time aggregation should only affect Tables 2-3 through the calculation of the standard errors. Using their suggested correction makes little difference to the t-statistics reported in either Table. (Indeed, $\mathrm{Corr}(e_t, e_{t-k})$ is small for all $k \leq 10$.)

Table 2

The Equity Premium Puzzle

α	$\bar{e}$	t-stat
0.0	0.0594	3.345
0.5	0.0577	3.260
1.0	0.0560	3.173
1.5	0.0544	3.082
2.0	0.0528	2.987
2.5	0.0512	2.890
3.0	0.0496	2.790
3.5	0.0480	2.688
4.0	0.0464	2.584
4.5	0.0449	2.478
5.0	0.0433	2.370
5.5	0.0418	2.262
6.0	0.0403	2.153
6.5	0.0390	2.044
7.0	0.0372	1.934
7.5	0.0357	1.824
8.0	0.0341	1.715
8.5	0.0326	1.607
9.0	0.0310	1.501
9.5	0.0295	1.395
10.0	0.0279	1.291

In this table, $\bar{e}$ is the sample mean of $e_t = (C_{t+1}/C_t)^{-\alpha}(R^s_{t+1} - R^b_{t+1})$ and α is the coefficient of relative risk aversion. Standard errors are calculated using the implication of the theory that e_t is uncorrelated with e_{t-k} for all k; however, they are little changed by allowing e_t to be MA(1) instead. This latter approach to calculating standard errors allows for the possibility of time aggregation (see Hansen, Heaton, and Yaron (1994)).

and invest in stocks. Note that the marginal benefit of selling Treasury bills and buying stocks falls as the investor's degree of risk aversion rises[7].

Table 3 reports the sample mean of e^b_t for various values of α and $\beta = 0.99$. It shows that for β equal to 0.99 and α greater than 1, the representative household can gain at the margin by transferring consumption from the future to the present (that is, reducing its savings rate). This is the riskfree rate puzzle. High values of α imply that individuals view consumption in different periods as complementary. Such individuals find an upwardly sloped consumption profile less desirable because consumption is not the same in every period of life. As a result, an individual with a high value of α realizes more of a utility gain by reducing her savings rate. Note that the riskfree rate puzzle comes from the equity premium puzzle: there is a riskfree rate puzzle only if α is required to be larger than 1 so as to match up with the high equity premium[8].

It is possible to find parameter settings for the discount factor β and the coefficient of relative risk aversion α that exactly satisfy the sample versions of equations (2a') and (2b'). In particular, by

---

[7]This intuition becomes even more clear if one assumes that $\ln(C_{t+1}/C_t)$, $\ln(R^s_{t+1})$, and $\ln(R^b_{t+1})$ are jointly normally distributed. Then, equations (2a') and (2b') become:

(2a") $\quad E(r^s_t - r^b_t) - \alpha\mathrm{Cov}(g_t, r^s_t - r^b_t) + 0.5\mathrm{Var}(r^s_t) + 0.5\mathrm{Var}(r^b_t) = 0$

(2b") $\quad \ln(\beta) - \alpha E(g_t) + E(r^b_t) + 0.5\{\alpha^2\mathrm{Var}(g_t) - 2\alpha\mathrm{Cov}(g_t, r^b_t) + \mathrm{Var}(r^b_t)\} = 0$

where $(g_t, r^s_t, r^b_t) \equiv (\ln(C_t/C_{t-1}), \ln(R^s_t), \ln(R^b_t))$. In the data, $\mathrm{Cov}(g_t, r^b_t)$ is essentially zero. Because $E(r^s_t - r^b_t)$ is so large, either α or $\mathrm{Cov}(g_t, r^s_t)$ must be large in order to satisfy (2a").

As with (2a') and (2b'), the sample analogs of (2a") and (2b") can only be satisfied by setting β equal to a value larger than 1 and α equal to a value greater than 15. See N. Gregory Mankiw and Stephen Zeldes (1991) for more details.

[8]See Edward Allen (1990), Cochrane and Hansen (1992), and Constantinides (1990) for similar presentations of the two puzzles.

Stephen Cecchetti, Pok-Sang Lam and Nelson Mark (1993) present the results of joint tests of (2a',b'). I present separate t-tests because I think doing so provides more intuition about the failings of the representative agent model. Of course, there is nothing wrong statistically with doing two separate t-tests, each with size α, as long as one keeps in mind that the size of the overall test could be as low as zero or it could be as high as 2α. (There may be a loss of power associated with using sequential t-tests but power doesn't seem to be an issue in tests of representative agent models!)

Table 3

The Riskfree Rate Puzzle

α	$\overline{e}$	t-stat
0.0	0.0033	0.0567
0.5	-0.0081	-1.296
1.0	-0.0162	-2.233
1.5	-0.0239	-2.790
2.0	-0.0313	-3.100
2.5	-0.0382	-3.263
3.0	-0.0448	-3.339
3.5	-0.0510	-3.360
4.0	-0.0569	-3.348
4.5	-0.0624	-3.312
5.0	-0.0675	-3.259
5.5	-0.0723	-3.195
6.0	-0.0768	-3.123
6.5	-0.0808	-3.043
7.0	-0.0846	-2.959
7.5	-0.0880	-2.871
8.0	-0.0910	-2.779
8.5	-0.0937	-2.685
9.0	-0.0960	-2.590
9.5	-0.0980	-2.492
10.0	-0.0997	-2.394

In this table, $\overline{e}$ is the sample mean of $e_t = \beta(c_{t+1}/c_t)^{-\alpha}R^b_{t+1} - 1$ and α is the coefficient of relative risk aversion. The discount factor β is set equal to 0.99. The standard errors are calculated using the implication of the theory that e_t is uncorrelated with e_{t-k} for all k; however, they are little changed by allowing e_t to be MA(1) instead. This latter approach to calculating standard errors allows for the possibility of time aggregation (see Hansen, Heaton, and Yaron (1994)).

setting $\alpha = 17.95$, it is possible to satisfy the equity premium first order condition (2a'); by also setting $\beta = 1.08$, it is possible to satisfy the riskfree rate first order condition (2b'). (See Kocherlakota (1990a) for a demonstration of how competitive equilibria can exist even when $\beta > 1$.) It is necessary to set α to such a large value because stocks offer a huge premium over bonds, and aggregate consumption growth does not covary greatly with stock returns; hence, the representative investor can only be marginally indifferent between stocks and bonds if he is highly averse to consumption risk. Similarly, the high consumption growth enjoyed by the United States since 1890 can only be consistent with high values of α and the riskfree rate if the representative investor is so patient that his 'discount' factor β is greater than one.

Thus, the equity premium and riskfree rate puzzles are solely a product of the parametric restrictions imposed by Mehra and Prescott on the discount factor β and the coefficient of relative risk aversion α. Given this, it is important to understand the sources of these restrictions. The restriction that the discount factor β is less than one emerges from introspection: most economists, including Mehra and Prescott, believe that an individual faced with a constant consumption stream would, on the margin, like to transfer some consumption from the future to the present.

The restriction that α should be less than ten is more controversial. Mehra and Prescott (1985) quote several microeconometric estimates that bound α from above by three. Unfortunately, the only estimate that they cite from financial market data has been shown to be severely biased downwards (Kocherlakota (1990c)). In terms of introspection, it has been argued by Mankiw and Zeldes (1991) that an individual with a coefficient of relative risk aversion above ten would be willing to pay unrealistically large amounts to avoid bets. However, they only consider extremely large bets (ones with potential losses of 50% of the gambler's wealth). In contrast, Shmuel Kandel and Robert Stambaugh (1991) show that even values of α as high as thirty imply quite reasonable behavior when the bet involves a maximal potential loss of around 1% of the gambler's wealth.

Because of the arguments offered by Kandel and Stambaugh (1991) and Kocherlakota (1990c), some economists[9] believe that there is no equity premium puzzle: individuals are more risk averse than we thought, and this high degree of risk aversion is reflected in the spread between stocks and bonds. However, it is clear to me from conversations and from knowledge of their work that a vast majority of economists believe that values for α above ten (or, for that matter, above five) imply highly implausible behavior on the part of individuals. (For example, Mehra and Prescott (1985, 1988) clearly chose an upper bound as large as ten merely as a rhetorical flourish.) D. Lucas (1994, p. 335) claims that any proposed solution that "does not explain the premium for $\alpha \leq 2.5$ ... is ... likely to be widely viewed as a resolution that depends on a high degree of risk aversion." She is probably not exaggerating the current state of professional opinion.

Tables 2 and 3 show that any model that leads to (2a',b') will generate the two puzzles. This helps make clear what the puzzles are *not* about. For example, while Mehra and Prescott ignore sampling error in much of their discussion, Tables 2 and 3 do not. This allays the concerns expressed by Allan Gregory and Gregor Smith (1991) and Cecchetti, Lam and Mark (1993) that the puzzles could simply be a result of sampling error. Similarly, Mehra and Prescott assume that the growth rate of the dividend to the S & P 500 is perfectly correlated with the growth rate of per capita consumption. Simon Benninga and Aris Protopapadakis (1990) argue that this might be responsible for the conflict between the model and the data. However, in Tables 2 and 3, the equity premium and riskfree rate puzzles are still present when the representative investor is faced with the real return to the S & P 500 itself, and not some imaginary portfolio with a dividend that is perfectly correlated with consumption.

Finally, we can see that the exact nature of the process generating consumption growth is irrelevant. For example, Cecchetti, Lam and Mark (1993), and Kandel and Stambaugh (1990) have both

[9]See, among others, Craig Burnside (1994), John Campbell and Cochrane (1994), Cecchetti and Mark (1990), Cecchetti, Lam and Mark (1993), and Hansen, Thomas Sargent and Thomas Tallarini (1994).

proposed that allowing consumption growth to follow a Markov *switching* process (as described by James Hamilton (1989)) might explain the two puzzles. Tables 2 and 3, though, show that this conjecture is erroneous[10]: the standard t-test is asymptotically valid when consumption growth follows any stationary and ergodic process, including one that is Markov switching[11].

Thus, Tables 2 and 3 show that there are really only three crucial assumptions generating the two puzzles. First, asset markets are complete. Second, asset markets are frictionless. (As discussed above, these two assumptions imply that there is a representative consumer.) Finally, the representative consumer's utility function has the power form assumed by Mehra and Prescott (1985) and satisfies their parametric restrictions on the discount factor and coefficient of relative risk aversion. Any model that seeks to resolve the two puzzles must weaken at least one of these three assumptions.

III. PREFERENCE MODIFICATIONS

The paradigm of complete and frictionless asset markets underlies some of the fundamental insights in both finance and economics. (For example, the Modigliani-Miller Theorems and the appeal of profit maximization as an objective for firms both depend on the completeness of asset markets).

---

[10]See Abel (1992) for another argument along these lines.

[11]Some readers might worry that the justification for Tables 2-3 is entirely asymptotic. I am sympathetic to this concern. I have run a Monte Carlo experiment in which the pricing errors e_t^s and e_t^b are assumed to be i.i.d. over time and have a joint distribution equal to their empirical joint distribution. In this setting, the Monte Carlo results show that the asymptotic distribution of the t-statistic and its distribution with 100 observations are about the same.

However, there is a limit to the generality of Tables 2 and 3. One could posit, as in Kocherlakota (1994) and Thomas Rietz (1988), that there is a low probability of consumption falling by an amount that has never been seen in United States data. With power utility, investors are very averse to these large falls in consumption, even if they are highly unlikely. Hence, investors believe that stocks are riskier and bonds are more valuable than appears to be true when one looks at information in the ninety year United States sample; the t-statistics are biased upwards. While this explanation is consistent with the facts, it has the defect of being largely nontestable using only returns and consumption data. (See Mehra and Prescott (1988) for another critique of this argument.)

Hence, it is important to see whether it is possible to explain the two puzzles without abandoning this useful framework. To do so, we must consider possible alterations in the preferences of the representative individual; in this section, I consider three different modifications to the preferences (1): generalized expected utility, habit formation, and relative consumption effects.

1. Generalized Expected Utility

In their (1989) and (1991) articles[12], Larry Epstein and Stanley Zin describe a generalization of the "standard" preference class (1). In these Generalized Expected Utility preferences, the period t utility U_t received by a consumer from a stream of consumption is described recursively using the formula:

$$U_t = \{c_t^{1-\rho} + \beta\{E_t U_{t+1}^{1-\alpha}\}^{(1-\rho)/(1-\alpha)}\}^{1/(1-\rho)} \tag{3}$$

Thus, utility today is a constant elasticity function of current consumption and future utility. Note that (1) can be obtained as a special case of these preferences by setting $\alpha = \rho$.

Epstein and Zin point out a crucial attribute of these preferences. In the preferences (1), the coefficient of relative risk aversion α is constrained to be equal to the reciprocal of the elasticity of intertemporal substitution. Thus, highly risk averse consumers must view consumption in different time periods as being highly complementary. This is not true of the GEU preferences (3). In this utility function, the degree of risk aversion of the consumer is governed by α while the elasticity of intertemporal substitution is governed by $1/\rho$. Epstein and Zin (1991) argue that disentangling risk

[12]Weil (1989), Kandel and Stambaugh (1991), and Hansen, Sargent and Tallarini (1994) provide complementary analyses of the implications of this type of generalized expected utility for asset pricing.

There are other ways to generalize expected utility so as to generate interesting implications for asset prices - see Epstein and Tan Wang (1994) and Epstein and Zin (1990) for examples.

aversion and intertemporal substitution in this fashion may help explain various aspects of asset pricing behavior that appear anomalous in the context of the preferences (1).

To see the usefulness of GEU preferences in understanding the equity premium puzzle and the riskfree rate puzzle, suppose there is a representative investor with preferences given by (3) who can invest in stocks and bonds. Then, the investor's optimal consumption profile must satisfy the two first order conditions:

$$\text{(4a)} \qquad E_t\{U_{t+1}^{\rho-\alpha}(C_{t+1}/C_t)^{-\rho}(R^s_{t+1} - R^b_{t+1})\} = 0$$

$$\text{(4b)} \qquad \beta E_t\{(E_t U_{t+1}^{1-\alpha})^{(\alpha-\rho)/(1-\alpha)} U_{t+1}^{\rho-\alpha}(C_{t+1}/C_t)^{-\rho} R^b_{t+1}\} = 1$$

Note that the marginal rates of substitution depend on a variable that is fundamentally unobservable: the period (t+1) level of utility of the representative agent (and the period t expectations of that utility). This unobservability makes it difficult to verify whether the data are consistent with these conditions.

To get around this problem, we can exploit the fact that it is difficult to predict future consumption growth using currently available information (see Hall (1978)). Hence, I assume in the discussion of GEU preferences that future consumption growth is statistically independent of all information available to the investor today[13]. Under this restriction, it is possible to show that utility in period (t+1) is a time and state invariant multiple of consumption in period (t+1). We can then apply the Law of Iterated Expectations to (4a) and (4b) to arrive at:

$$\text{(4a}') \qquad \beta E\{(C_{t+1}/C_t)^{-\alpha}(R^s_{t+1} - R^b_{t+1})\} = 0$$

[13]This assumption of serial independence is also employed by Abel (1990), Epstein and Zin (1990), and Campbell and Cochrane (1994). Kandel and Stambaugh (1991), Kocherlakota (1990b), and Weil (1989) have shown that the thrust of the discussion that follows is little affected by allowing for more realistic amounts of dependence in consumption growth.

(4b′) $$\beta[E(C_{t+1}/C_t)^{1-\alpha}]^{(\alpha-\rho)/(1-\alpha)}E\{(C_{t+1}/C_t)^{-\alpha}R^b{}_{t+1}\} = 1$$

See Kocherlakota (1990b) for a precise derivation of these first order conditions[14].

The first order condition (4a′) tells us that using GEU preferences will not change Table 2: for each value of α, the marginal utility gain of reducing bond holdings and investing in stocks is the same whether the preferences lie in the expected utility class (1) or in the generalized expected utility class (3). In contrast, Table 4 shows that it is possible to resolve the riskfree rate puzzle using the GEU preferences. In particular, for $\beta = 0.99$, and various values of α in the set (0, 18], Table 4 presents the value of ρ that exactly satisfies the first order condition (4b).

The intuition behind these results is simple. The main benefit of modelling preferences using (3) instead of (1) is that individual attitudes towards risk and growth are no longer governed by the same parameter. However, the equity premium puzzle arises only because of economists' prior beliefs about risk aversion; hence, this puzzle cannot be resolved by disentangling attitudes towards risk and growth. the other hand, the connection between risk aversion and intertemporal substitution in the standard preferences (1) *is* the essential element of the riskfree rate puzzle. To match the equity premium, risk

---

[14]Epstein and Zin (1991) derive a different representation for the first order conditions (4a) and (4b) of the representative investor:

(*) $$E_t\{(C_{t+1}/C_t)^{-\gamma\rho}(R^m_{t+1})^{\gamma-1}(R^s_{t+1} - R^b_{t+1})\} = 0$$
(**) $$E_t\{\beta^\gamma(C_{t+1}/C_t)^{-\gamma\rho}(R^m_{t+1})^{\gamma-1}R^b_{t+1})\} = 1$$

where $\gamma \equiv (1-\alpha)/(1-\rho)$. In (*) and (**), R^m_t is the gross real return to the representative investor's entire portfolio of assets (including human capital, housing, etc.). This representation has the desirable attribute of being valid regardless of the information set of the representative investor (so we don't have to assume that consumption growth from period t to period (t+1) is independent of all information available to the investor in period t). Unfortunately, the variable R^m_t is not observable. Epstein and Zin (1991) use the value-weighted return to the NYSE as a proxy, but of course this understates the true level of diversification of the representative investor and (potentially) greatly overstates the covariability of her marginal rate of substitution with the return to the stock market. As a result, using this proxy variable leads one to the spurious conclusion that GEU preferences can resolve the equity premium puzzle with low levels of risk aversion.

Table 4

Resolving the Riskfree Rate Puzzle with GEU Preferences

α	$1/\alpha$	$1/\rho$
0.500	2.000	23.93
1.000	1.000	14.91
1.500	0.6667	10.62
2.000	0.5000	8.126
2.500	0.4000	6.493
3.000	0.3333	5.345
3.500	0.2857	4.494
4.000	0.2500	3.840
4.500	0.2222	3.322
5.000	0.2000	2.903
5.500	0.1818	2.557
6.000	0.1667	2.267
6.500	0.1538	2.022
7.000	0.1429	1.811
7.500	0.1333	1.629
8.000	0.1250	1.469
8.500	0.1176	1.330
9.000	0.1111	1.206
9.500	0.1053	1.096
10.00	0.1000	0.9974
10.50	0.09524	0.9091
11.00	0.09091	0.8295
11.50	0.08696	0.7574
12.00	0.08333	0.6920
12.50	0.08000	0.6325
13.00	0.07692	0.5781
13.50	0.07407	0.5283
14.00	0.07143	0.4826
14.50	0.06897	0.4406
15.00	0.06667	0.4018
15.50	0.06452	0.3660
16.00	0.06250	0.3329
16.50	0.06061	0.3022
17.00	0.05882	0.2737
17.50	0.05714	0.2472
18.00	0.05556	0.2226

In this table, the first column is the coefficient of relative risk aversion α, the second column is the corresponding elasticity of intertemporal substitution if preferences are of the form (1), and the third column is the value of $1/\rho$ that satisfies the riskfree rate first order condition (4b′) given that preferences are of the form (3), $\beta = 0.99$ and α has the value specified in the first column.

aversion has to be high. But in the standard preferences (1), this forces intertemporal substitution to be low. The GEU preferences can explain the riskfree rate puzzle by allowing intertemporal substitution and risk aversion to be high simultaneously[15].

2. Habit Formation

We have seen that the riskfree rate puzzle is a consequence of the demand for savings being too low when individuals are highly risk averse and have preferences of the form (1). The GEU preference class represents one way to generate high savings demand when individuals are highly risk averse. There is another, perhaps more intuitive, approach. The standard preferences (1) assume that the level of consumption in period (t-1) does not affect the marginal utility of consumption in period t. It may be more natural to think that an individual who consumes a lot in period (t-1) will get used to that high level of consumption, and will more strongly desire consumption in period t; mathematically, the individual's marginal utility of consumption in period t is an *increasing* function of period (t-1) consumption.

This property of intertemporal preferences is termed *habit formation*. The basic implications of habit formation for asset returns (as explained by Constantinides (1990) and Heaton (1995)) can be captured in the following simple model. Suppose that there is a representative agent with preferences in period t represented by the utility function:

---

[15]Weil (1989) emphasizes that GEU preferences cannot explain the riskfree rate puzzle if one forces the coefficient of relative risk aversion to be "reasonable" (say, approximately 1). He shows that under this restriction, then the elasticity of intertemporal substitution must be very large to be consistent with the riskfree rate. (See Table 4 for a confirmation.) This large elasticity of intertemporal substitution appears to contradict the results of Hall (1988), who shows that the estimated slope coefficient is very small in a regression of consumption growth on expected real interest rates (although Weil (1990) points out that it is difficult to see a direct linkage between Hall's regression coefficients and the parameters that describe the GEU preferences).

(5) $$E_t \Sigma_{s=0}^{\infty} \beta^s (c_{t+s} - \lambda c_{t+s-1})^{1-\alpha}/(1-\alpha), \ \lambda > 0$$

Thus, the agent's momentary utility is a decreasing function of last period's consumption: it takes more consumption today to make him happy if he consumed more yesterday.

The individual's marginal utility with respect to period t consumption is given by the formula:

(6) $$MU_t = (c_t - \lambda c_{t-1})^{-\alpha} - \beta\lambda E_t(c_{t+1} - \lambda c_t)^{-\alpha}$$

Thus, the marginal utility of period t consumption is an increasing function of period (t-1) consumption. Note that there are two pieces to this formula for marginal utility. The first term captures the fact that if I buy a BMW rather than a Yugo today, then I am certainly better off today; however, the second piece models the notion that having the BMW "spoils" me and reduces my utility from all future car purchases.

The individual's optimal consumption portfolio must satisfy the two first order conditions that make him indifferent to buying or selling more stocks or bonds:

(7a) $$\beta E_t\{(MU_{t+1}/MU_t)(R^s_{t+1}-R^b_{t+1})\} = 0$$

(7b) $$\beta E_t\{(MU_{t+1}/MU_t)R^b_{t+1}\} = 1$$

where MU_t is defined as in (6). We can apply the Law of Iterated Expectations to obtain:

(7a′) $$\beta E\{(MU_{t+1}/MU_t)(R^s_{t+1}-R^b_{t+1})\} = 0$$

(7b′) $$\beta E\{(MU_{t+1}/MU_t)R^b_{t+1}\} = 1$$

The problem with trying to satisfy (7a′) and (7b′) is that MU_t depends on the investor's ability to predict future consumption growth. As with GEU preferences, this means that the formula for the marginal utility depends on the individual's (possibly unobservable) information. As in that context, it is convenient to assume that consumption growth from period t to period (t+1) is unpredictable. Then the ratio of marginal utilities is given by:

$$(8)\quad MU_{t+1}/MU_t = \{(\phi_{t+1}\phi_t-\lambda\phi_t)^{-\alpha} - \lambda\beta(\phi_{t+1}\phi_t)^{-\alpha}E(\phi-\lambda)^{-\alpha}\}/\{(\phi_t-\lambda)^{-\alpha} - \lambda\beta(\phi_t)^{-\alpha}E(\phi-\lambda)^{-\alpha}\}$$

where $\phi_t \equiv C_t/C_{t-1}$. This allows us to easily estimate MU_{t+1}/MU_t by using the sample mean to form an estimate of the unconditional expectations in the formula (8).

With the preference class (1), it is not possible to simultaneously satisfy equations (2a') and (2b') without driving β above one. This is not true of preferences that exhibit habit formation. For example, suppose we set $\beta = 0.99$. Unlike the preference class (1), it is then possible to find preference parameters to satisfy both (7a') and (7b'). In particular, if we set $\alpha = 15.384$ and $\lambda = 0.174$, then the sample analogs of both first order conditions are exactly satisfied in the data[16].

---

[16]Instead of (7a',b'), Wayne Ferson and Constantinides (1991) work with an alternative implication of (7a,b):

$$(*)\qquad E\{MU_tZ_t/(C_t-\lambda C_{t-1})^{-\alpha}\} = \beta E\{MU_{t+1}R_{t+1}Z_t/(C_t-\lambda C_{t-1})^{-\alpha}\}.$$

where R_{t+1} is the gross return to an arbitrary asset, and Z_t is an arbitrary variable observable to the econometrician and to the agent at time t. This condition has the desirable property that it doesn't involve terms that are unobservable to the econometrician. However, (*) is a different type of implication of optimal behavior from (4a',b') or (2a',b') because it is not just an averaged version of (7a,b) (even when Z_t is set equal to 1). In order to be consistent across preference orderings, I work with (7a',b').

In the context of a calibrated general equilibrium model, Constantinides (1990) generates a large equity premium and a low riskfree rate by assuming that the representative agent's utility function has a large value of λ and a low value of α. However, it is important to keep in mind that when λ is large, the individual requires a large level of consumption in order to just survive; he will pay a lot to avoid small consumption gambles even if α is low. Thus, Constantinides' proposed resolution of the puzzles requires individuals to be very averse to consumption risk (although, as he shows, not to wealth risk).

There are two implications of this result. First, habit formation does not resolve the equity premium puzzle: it is still true that the representative investor is indifferent between stocks and bonds only if she is highly averse to consumption risk. On the other hand, habit formation does help resolve the riskfree rate puzzle - it is possible to satisfy (7a′) and (7b′) without setting β larger than one. The intuition behind this finding is simple. For any given level of current consumption, the individual knows that his desire for consumption will be higher in the future because consumption is habit-forming; hence, the individual's demand for savings increases relative to the preference specification (1).

3. Relative Consumption: 'Keeping up with the Joneses'

The standard preferences (1) assume that individuals derive utility only from their own consumption. Suppose, though, that as James Duesenberry (1952) posits, an individual's utility is a function not just of his own consumption but of societal levels of consumption. Then, his investment decisions will be affected not just by his attitudes towards his own consumption risk, but also by his attitudes towards variability in societal consumption. Following up on this intuition, Abel (1990) and Jordi Gali (1994) examine the asset pricing implications of various classes of preferences in which the utility an individual derives from a given amount of consumption depends on per capita consumption[17].

I will work with a model that combines features of Abel's and Gali's formulations. Suppose the representative individual has preferences in period t:

$$(9) \qquad E_t \sum_{s=0}^{\infty} \beta^s c_{t+s}^{1-\alpha} C_{t+s}^{\gamma} C_{t+s-1}^{\lambda}/(1-\alpha), \ \alpha \geq 0$$

where c_{t+s} is the individual's level of consumption in period (t+s) while C_{t+s} is the level of per capita

[17]See Campbell and Cochrane (1994) and James Nason (1988) for models which feature both relative consumption effects and time-varying risk aversion.

consumption in period (t+s) (in equilibrium, of course, the two are the same). Thus, the individual derives utility from how well she is doing today *relative* to how well the average person is doing today and how well the average person did last period. If the individual is a jealous sort, then it is natural to think of γ and λ as being negative: the individual is unhappy when others are doing well. If the individual is patriotic, then it is natural to think of γ and λ as being positive: the individual is happy when per capita consumption for the nation is high.

The representative individual treats per capita consumption as exogenous when choosing how much to invest in stocks and bonds. Hence, in equilibrium, the individual's first order conditions take the form:

(10a) $$E_t\{(C_{t+1}/C_t)^{\gamma-\alpha}(C_t/C_{t-1})^{\lambda}(R^s{}_{t+1} - R^b_{t+1})\} = 0$$

(10b) $$\beta E_t\{(C_{t+1}/C_t)^{\gamma-\alpha}(C_t/C_{t-1})^{\lambda}R^b{}_{t+1}\} = 1$$

We can apply the law of iterated expectations to derive the two conditions:

(10a') $$E\{(C_{t+1}/C_t)^{\gamma-\alpha}(C_t/C_{t-1})^{\lambda}(R^s{}_{t+1} - R^b_{t+1})\} = 0$$

(10b') $$\beta E\{(C_{t+1}/C_t)^{\gamma-\alpha}(C_t/C_{t-1})^{\lambda}R^b{}_{t+1}\} = 1$$

For any specification of the discount factor β and the coefficient of relative risk aversion α, it is possible to find settings for γ and λ such that two puzzles are resolved in that the sample analogs of (10a') and (10b') are satisfied. For example, suppose $\beta = 0.99$. Then, if $(\alpha-\gamma) = 19.280$, and $\lambda = 3.813$, the sample analogs of (10a') and (10b') are satisfied exactly.

This model offers the following explanation for the seemingly large equity premium. When γ is large in absolute value, an individual's marginal utility of own consumption is highly sensitive to fluctuations in per capita consumption and therefore strongly negatively related to stock returns. Thus, even if α is small so that the "representative" investor is not all that averse to individual consumption risk, she does not find stocks attractive because she is highly averse to *per capita* consumption risk.

However, the presence of contemporaneous per capita consumption in the representative investor's utility function does not help explain the riskfree rate puzzle: if λ is set equal to zero, (10a′) and (10b′) become equivalent to (2a′) and (2b′). We know from our analysis of those equations that if $\lambda = 0$, it is necessary to drive β larger than one in order to satisfy (10b′) given that (10a′) is satisfied. The positive effect of *lagged* per capita consumption on current marginal utility increases the individual's demand for savings and allows (10a′) to be satisfied with a relatively low value of β.

4. Summary

Mehra and Prescott (1985) require the preferences of the representative individual to lie in the "standard" class (1). We have seen that these preferences can only be made consistent with the large equity premium if the coefficient of relative risk aversion is pushed near twenty. Given the low level of the average Treasury bill return, individuals who see consumption in different periods as so complementary will only be happy with the steep consumption profile seen in United States data if their 'discount' factor is above one.

Several researchers have explored the consequences of using broader classes of preferences. Broadly speaking, there are two lessons from this type of sensitivity analysis. First, the riskfree rate puzzle can be resolved as long as the link between individual attitudes towards risk and growth contained in the standard preferences (1) is broken. Second, the equity premium puzzle is much more robust: individuals must either be highly averse to their own consumption risk or to per capita consumption risk if they are to be marginally indifferent between investing in stocks or bonds.

It seems that any resolution to the equity premium puzzle in the context of a representative agent model will have to assume that the agent is highly averse to consumption risk. Per capita consumption is very smooth, and therefore does not covary greatly with stock returns. Yet people continue to demand a high expected return for stocks relative to bonds. The only possible conclusion is that individuals are extremely averse to any marginal variation in consumption (either their own or societal).

IV. INCOMPLETE MARKETS AND TRADING COSTS

In this section, I examine the ability of models with incomplete markets and various sorts of trading frictions to explain the asset returns data[18]. Unlike the previous section, which focused solely on the decision problem of a "representative" agent, the analysis will (necessarily) be fully general equilibrium in nature. Because of the complexity of these dynamic general equilibrium models, their implications are generated numerically; I find that this makes any intuition more speculative in nature.

Throughout this section, all investors are assumed to have identical preferences of the form (1). The "game" in this literature is to explain the existence of the puzzles solely through the presence of incompleteness and trading costs that Mehra and Prescott assume away (just as in Section III, the "game" was to explain the puzzle without allowing such frictions).

1. Incomplete Markets

Underlying the Mehra-Prescott model is the presumption that the behavior of per capita consumption growth is a good guide to the behavior of individual consumption growth. We have seen that this belief is warranted if asset markets are *complete* so that individuals can write contracts against any contingency: individuals will use the financial markets to diversify away any idiosyncratic differences

---

[18]Throughout my discussion, I ignore the existence of fiat money. Pamela Labadie (1989) and Alberto Giovannini and Labadie (1991) show that motivating a demand for money via a cash-in-advance constraint does not significantly alter the asset pricing implications of the Mehra-Prescott (1985) model. However, a richer model of money demand might have more dramatic effects on asset prices.

in their consumption streams. As a result, their consumption streams will look similar to each other and to per capita consumption.

However, in reality, it does not appear easy for individuals to directly insure themselves against all possible fluctuations in their consumption streams. (For example, it is hard to directly insure oneself against fluctuations in labor income.) For this reason, most economists believe that insurance markets are *incomplete*. Intuitively, in the absence of these kinds of markets, individual consumption growth will feature risk not present in per capita consumption growth and so individual consumption growth will be more variable than per capita consumption growth. As a result, models with incomplete markets offer the hope that while the covariance of per capita consumption growth with stock returns is small, individual consumption growth may covary enough with stock returns to explain the equity premium.

a. Dynamic Self-Insurance and the Riskfree Rate

The intuitive appeal of incomplete markets for explaining asset returns data is supported by the work of Weil (1992). He studies a two period model in which financial markets are incomplete. He shows that if individual preferences exhibit prudence (that is, convex marginal utility - see Miles Kimball (1990)), the extra variability in individual consumption growth induced by the absence of markets helps resolve the riskfree rate puzzle. Without complete markets, individuals must save more in order to self-insure against the randomness in their consumption streams; the extra demand for savings drives down the riskfree rate. Weil also shows that if individuals exhibit not just prudence but decreasing absolute prudence (see Kimball (1990)), the additional variability in consumption growth induced by market incompleteness also helps to explain the equity premium puzzle. The extra riskiness in individual consumption makes stocks seem less attractive to an individual investor than they would to a "representative" consumer[19].

Unfortunately, two period models cannot tell the full story because they abstract from the use of *dynamic* trading as a form of insurance against risk. For example, suppose my salary next year is equally likely to be $40000 or $50000. If I die at the end of the year and don't care about my descendants, then the variability in income has to be fully reflected in my expenditure patterns. This is the behavior that is captured by two period models. On the other hand, if I know I will live for many more years, then I need not absorb my income risk fully into current consumption - I can partially offset it by reducing my savings when my income is high and increasing my savings when my income is low.

To understand the quantitative implications of dynamic self-insurance for the riskfree rate, it is best to consider a simple model of asset trade in which there are a large number (in fact, a continuum) of *ex-ante* identical infinitely-lived consumers. Each of the consumers has a random labor income and their labor incomes are independent of one another (so there is no variability in per capita income). The consumers cannot directly write insurance contracts against the variability in their individual income streams. In this sense, financial markets are incomplete. However, the individuals are able to make riskfree loans to one another, although they cannot borrow more than B in any period.

For now, I want to separate the effects of binding borrowing constraints from the effects of incomplete markets. To eliminate the possibility that the debt ceiling B is ever a binding constraint, I assume throughout the following discussion that B is larger than y_{min}/r, where y_{min} is the lowest possible realization of income. No individual will borrow more than y_{min}/r; doing so would mean that with a sufficiently long run of bad income realizations, the individual would be forced into default, which is assumed to be infinitely costly (see Aiyagari (1994)). Market clearing dictates that for every lender, there

[19]Mankiw (1986) shows that if the conditional variability of individual income shocks is higher when the aggregate state of the economy is lower, then prudence alone is sufficient to generate a higher equity premium in the incomplete markets environment.

must be a borrower; hence, y_{min} must be larger than 0 if there is to be trade in equilibrium[20].

As in the two period context, an individual in this economy faces the possibility of an uninsurable consumption fall. The desire to guard against this possibility (at least on the margin) generates a greater demand to transfer resources from the present to the future than in a complete markets environment. This extra demand for saving forces the equilibrium interest rate below the complete markets interest rate[21].

However, the ability to *dynamically* self-insure in the infinite horizon setting means that the probability of an uninsurable consumption fall is much smaller than in the two period context. In equilibrium, the "typical" individual has no savings (because net assetholdings are zero), but does have a "line of credit" y_{min}/r. The individual can buffer any short-lived falls in consumption by borrowing; his line of credit can only be exhausted by a relatively unlikely long "run" of bad income realizations. (Equivalently, in a stationary equilibrium, few individuals are near their credit limit at any given point in time.) Thus, the extra demand for savings generated by the absence of insurance markets will generally be smaller in the infinite horizon economy than in a two period model.

This reasoning suggests that in the infinite horizon setting, the absence of income insurance markets may have little impact on the interest rate. The numerical work of Mark Huggett (1993) and Heaton and D. Lucas (1995b) confirms this intuition. They examine infinite horizon economies which

---

[20]Christopher Carroll (1992) argues using data from the Panel Study in Income Dynamics that y_{min} equals zero. In this case, no individual can ever borrow (because their debt ceiling equals zero). There can only be trade of riskfree bonds if there is an outside supply of them (say, by the government). Then, the riskfree rate puzzle becomes an issue of how *low* the amount of outside debt must be in order to generate a plausibly low value for the riskfree rate. There is no clean answer to this question in the existing literature.

[21]Marilda Sotomayor (1984) proves that the individual demand for assetholdings and for consumption grows without bound over time if the interest rate equals the rate of time preference (the complete markets interest rate). (Her result is valid without the assumption of convex marginal utility as long as the consumption set is bounded from below.) This implies that the equilibrium interest rate must fall below the rate of time preference in order to deter individuals from saving so much. See Zeldes (1989b) and Angus Deaton (1992) for numerical work along these lines in a finite horizon setting.

are calibrated to accord with aspects of United States data on individual labor income and find that the difference between the incomplete markets interest rate and the complete markets interest rate is small[22].

The above discussion assumes that individual income shocks die out eventually (that is, the shocks are stationary). Constantinides and Darrell Duffie (1995) point out that the story is very different when shocks to individual labor income are permanent instead of transitory. Under the assumption of permanence, dynamic self-insurance can play no role: income shocks must be fully absorbed into consumption. For example, if my salary falls by $10,000 permanently, then the only way to keep my consumption smooth over time is reduce it by the same $10,000 in every period: I cannot just temporarily run down my savings account to smooth consumption. Since income shocks must be fully absorbed into consumption, the results from the two period model become much more relevant. Indeed, Constantinides and Duffie (1995) show that the absence of labor income insurance markets, combined with the permanence of labor income shocks, has the potential to generate a riskfree rate that may be much lower than the complete markets riskfree rate.

The issue, then, becomes an empirical one: how persistent (and variable) are otherwise undiversifiable shocks to individual income? This is a difficult matter to sort out because of the paucity

---

[22]In fact, they obtain these results in economies in which some fraction of the agents face binding borrowing constraints in equilibrium. Relaxing these borrowing constraints or adding outside debt will serve to further increase the equilibrium interest rate towards the complete markets rate (Aiyagari (1994)).

There is a caveat associated with the results of Huggett (1993) and Heaton and D. Lucas (1995b) (and with most of the other numerical work described later in this paper). For computational reasons, these papers assume that the process generating individual income is discrete. This discretization may impose an artificially high lower bound on income - that is, an artificially high value of y_{min}. Thus, Heaton and D. Lucas (1995b) assume that income cannot fall below 75% of its mean value. Huggett (1993) is more conservative in assuming that income cannot fall below about 10% of its mean value. In both cases, finer and more accurate discretizations may drive y_{min} lower.

This technical detail may have important consequences for the numerical results because no individual can ever borrow more than y_{min}/r. As y_{min} falls, there is an increased probability of a fall in consumption that cannot be "buffered" by using accumulated savings or by borrowing, and the downward pressure on the equilibrium interest rate rises. In fact, as y_{min} goes to zero, the equilibrium interest rate must go to zero (if there is no outside debt).

of time series evidence available. However, Heaton and D. Lucas (1995a) use data from the Panel Study of Income Dynamics and estimate the autocorrelation of idiosyncratic[23] labor income shocks to be around 0.5. When they examine the predictions of the above model using this autocorrelation for labor income (or even an autocorrelation as high as 0.8), they find that the equilibrium interest rate is close to the complete markets rate.

b. Dynamic Self-Insurance and the Equity Premium Puzzle

The above analysis shows that as long as most individuals are sufficiently far from the debt ceiling y_{min}/r, dynamic self-insurance implies that the equilibrium interest rate in incomplete markets models is well-approximated by the complete markets interest rate. In this section, I claim that a similar conclusion can be reached about the equity premium as long as the interest rate and the persistence of shocks are both sufficiently low.

The following argument is a crude intuitive justification for this claim, based largely on the Permanent Income Hypothesis. Consider an infinitely lived individual who faces a constant interest rate r and who faces income shocks that have a positive autocorrelation equal to ρ; if $\rho = 1$, the shocks are permanent while if $\rho < 1$, their effect eventually dies out. Financial markets are incomplete in the sense that the individual cannot explicitly insure against the income shocks. Because of concave utility, the individual wants to smooth consumption as much as possible.

Suppose the individual's income is surprisingly higher in some period by δ dollars. Because of the autocorrelation in the income shocks, the present value of his income rises by:

---

[23]What matters is the persistence of the *undiversifiable* component of labor income. Heaton and Lucas (1995a) implicitly assume that individuals can hedge any labor income risk that is correlated with aggregate consumption. Hence, they obtain their autocorrelation estimate of 0.5 only after first controlling for aggregate effects upon individual labor income (this may explain why their estimate of persistence is so much lower than that of Glenn Hubbard, Jonathan Skinner, and Zeldes (1994), who control for aggregate effects only through the use of year dummies).

$$\delta + \rho\delta/(1+r) + \rho^2\delta/(1+r)^2 + \ldots.$$

$$= \delta/(1 - \rho/(1+r))$$

$$= \delta(1+r)/(1+r-\rho).$$

The individual wants to smooth consumption. Hence, he increases consumption in every period of life by the annuitized value of the increase in his wealth. A simple present value calculation implies that his consumption rises by $\delta r/(1+r-\rho)$ dollars.

If the individual were able to fully insure against all income shocks, then his consumption would not change at all in response to the surprise movement in his income process. Thus, his ability to self-insure against income shocks using only savings can be measured by how close $r/(1+r-\rho)$ is to zero. For example, if $\rho = 1$, and all shocks are permanent, then $r/(1+r-\rho)$ is equal to one; the income shock is fully absorbed into consumption, which is very different from the implications of the full insurance model. In contrast, if $\rho < 1$, then as r nears zero, the implications of the full insurance model and the incomplete markets models for movements in individual consumption become about the same. In the United States annual data, we have seen that r is around 1%; hence, even if ρ is as high as 0.75, consumption is fairly insensitive to surprise movements in labor income. (Recall that Heaton and D. Lucas (1995a) estimate the autocorrelation of undiversifiable income shocks to be 0.5.)

This intuitive argument is borne out by the numerical work of Christopher Telmer (1993) and D. Lucas (1994)[24]. These papers examine the quantitative predictions of dynamic incomplete markets models, calibrated using individual income data, for the equity premium. They find that dynamic self-insurance allows individuals to closely approximate the allocations in a complete markets environment.

---

[24]See also Heaton and D. Lucas (1995a,b), Wouter den Haan (1994) and Albert Marcet and Singleton (1991).

The equilibrium asset prices are therefore similar to the predictions of the Mehra-Prescott (1985) model.

A critical assumption underlying this numerical work is that income shocks have an autocorrelation less than one. Constantinides and Duffie (1995) show that if labor income shocks are instead permanent, then it is possible for incomplete markets models to explain the large size of the equity premium. As in the case of the riskfree rate puzzle, the failure of dynamic self-insurance when $\rho = 1$ means that the model's implications resemble those of a two period framework.

To sum up: the assumption that financial markets are complete strikes many economists as *prima facie* ridiculous, and it is tempting to conclude that it is responsible for the failure of the Mehra-Prescott model to explain the United States data on average stock and bond returns. However, as long as investors can costlessly trade *any* financial asset (for example, riskfree loans) over time, they can use the accumulated stock of the asset to self-insure against idiosyncratic risk (for example, shocks to individual income.) The above discussion shows that given this ability to self-insure, the behavior of the riskfree rate and the equity premium are largely unaffected by the absence of markets as long as idiosyncratic shocks are not highly persistent.

2. Trading Costs

In the incomplete and complete markets models described above, it is assumed that individuals can costlessly trade any amount of the available securities. This assumption is unrealistic in ways that might matter for the two puzzles. For example, in order to take advantage of the possible utility gains associated with the large premium offered by equity, individuals have to reduce their holdings of Treasury bills and buy more stocks. The costs of making these trades may wipe out their apparent utility benefits. In this subsection, I consider the implications for asset returns of various types of trading costs.

a. Borrowing and Short Sales Constraints

Consider the familiar two period model of borrowing and lending in most undergraduate microeconomic textbooks. In this model, the individual's budget set of consumption choices (c_1, c_2) is written as:

$$c_1 \leq y_1 + b_1$$

$$c_2 + b_1(1+r) \leq y_2$$

$$c_1 \geq 0, c_2 \geq 0$$

where b_1 is the level of borrowing in period 1. The budget set allows the individual to borrow up to the present value of his period two income.

Many economists believe that this model of borrowing and lending ignores an important feature of reality: because of enforcement and adverse selection problems, individuals are generally not able to fully capitalize their future labor income. One way to capture this view of the world is to add a constraint to the individual's decision problem, $b_1 \leq B$, where B is the limit on how much the individual can borrow. If B is less than $y_2/(1+r)$, the constraint on b_1 may be binding because the individual may not be able to transfer as much income from period two to period one as he desires.

The restriction $b_1 \leq B$ is called a *borrowing constraint*. (*Short sales* constraints are similar restrictions on the trade of stocks.) The law of supply and demand immediately implies that the equilibrium interest rate is generally lower if many individuals face a binding borrowing constraint than if few individuals do. Intuitively, "tighter" borrowing constraints exogenously reduce the size of the borrowing side of the market; in order to clear markets, interest rates must fall in order to shrink the size of the lending side of the market.

In keeping with this intuition, the work of Huggett (1993) and Heaton and D. Lucas (1995a,b) demonstrates numerically that if a sizeable fraction of individuals face borrowing constraints, then the riskfree rate may be substantially lower than the predictions of the representative agent model. Using data from the Panel Study on Income Dynamics, Zeldes (1989a) documents that the large number of individuals with low levels of assetholdings appear to face such constraints.

However, Heaton and D. Lucas (1995a,b) show that borrowing and short sales constraints do not appear to have much impact on the size of the equity premium. There is a sound intuition behind these results: an individual who is constrained in the stock market generally must also be constrained in the bond market (or vice-versa). Otherwise, the individual could loosen the constraint in the bond market by shifting resources to it from the stock market. Hence, just as the average riskfree rate must fall to clear the bond market when many individuals are constrained, so must the average stock return fall in order to clear the constrained stock market[25].

b. Transaction Costs

Individuals who try to engage in asset trade quickly find that they face all sorts of transaction costs. (These include informational costs, brokerage fees, load fees, the bid-ask spread, and taxes.) To understand how transaction costs affect the pricing of securities, consider an investor in a riskfree world who can invest in two different securities, stocks and bonds. Stocks pay a constant dividend equal to d, and have a constant price equal to p. Because stocks are a perpetuity, their rate of return r_s, ignoring transaction costs, equals d/p. Bonds are short-lived and pay a constant interest rate r_b. There is no cost associated with trading bonds. However, stocks are costly to trade in that a buyer of stocks must incur a cost $p\tau$, in addition to the price p, to buy a share.

[25]Following Hua He and David Modest (1995), Cochrane and Hansen (1992) show that this argument is exactly correct when individuals face a *market wealth* constraint that restricts the total value of their asset portfolio not to fall below some exogenously specified (negative) number.

In equilibrium, the cost of buying a share of stock cannot be smaller than the present value of the benefits of owning that share for N periods (or there would be excess demand for stocks because everyone would want to buy shares). This statement can be expressed mathematically as:

$$p(1+\tau) \geq d/(1+r_b) + d/(1+r_b)^2 + \ldots + d/(1+r_b)^N + p/(1+r_b)^N$$

This present value expression can be rewritten as:

$$r_s \equiv d/p \leq (1+\tau)r_b/\{1-(1+r_b)^{-N}\}$$

Note that the right hand side gets smaller as N grows large. Intuitively, as the investor holds the stock for a long period of time, he is able to more fully amortize the initial trading cost; the individual is therefore only prevented from making arbitrage profits if stock returns are extremely close to those of bonds. Indeed, when N equals infinity, the upper bound on the premium paid by stocks is very tight:

$$r_s - r_b \leq \tau r_b$$

Thus, if $r_b = 1\%$, and investors can buy stocks and hold them forever, then r_s can be as large as 7% in equilibrium only if τ is greater than 600%!

In the above riskfree model, an investor could buy and hold a stock for an infinite number of periods. However, if the investor instead faces income risk which cannot be directly insured, she is not able to plan on being able to hold stocks forever. The reason is simple: if she is ever pushed up against a debt ceiling, so she cannot borrow to fully offset bad income shocks, she will have to sell the stocks in order to smooth her consumption. This cap on the investor's "holding period" generated by potential runs of bad luck raises the possibility that in models with missing income insurance markets, transaction costs could have big effects on asset return spreads.

However, from our discussion of calibrated incomplete markets models, we know that an infinitely-lived individual can do a pretty good job of smoothing consumption by simply buying and selling the asset that is cheaper to trade. Hence, the agent faces only a low probability of needing to sell stocks in order to smooth consumption. She can consequently contemplate making arbitrage profits by buying and holding stocks for a long period of time; as we have seen, this means that stocks cannot pay a much higher return than bonds in equilibrium.

Thus, as Aiyagari and Mark Gertler (1991) and Heaton and D. Lucas (1995a) find, the only way to explain the equity premium using transaction costs is to assert that there are significant differences in trading costs across the stock and bond markets. In my view, there is little evidence to support this proposition at present, although both Aiyagari and Gertler (1991) and Heaton and D. Lucas (1995b) make some attempts in this direction. To be considered as the leading explanation of the puzzle, more needs to be done to document the sizes and sources of trading costs (especially for pension funds and other institutional investors that operate on behalf of stockholders)[26].

c. Market Segmentation

A key aspect of the Mehra-Prescott (1985) model is that it treats per capita consumption growth as being a good proxy for individual consumption growth. Yet, Mankiw and Zeldes (1991) and Michael Haliassos and Carol Bertaut (1991) document that only about 30% of individuals in the United States own stocks (either directly or through defined contribution pension funds). This suggests the possibility of *market segmentation*: that for whatever reason, only a subset of investors are actively involved in asset trade. Campbell (1993) points out that it is the per capita consumption growth of the active traders that should be substituted into first order conditions like (2a) and (2b), not overall per capita consumption growth as Mehra and Prescott (1985) use.

However, Mankiw and Zeldes (1991) provide direct evidence that market segmentation alone

[26]In their partial equilibrium settings, Erzo Luttmer (1994) and He and Modest (1995) assume that all investors fully turn over their portfolios relatively frequently (once a month or once a quarter) and all at the same time. This assumption allows them to explain the equity premium with relatively small transaction costs; however, it seems difficult to build a general equilibrium model in which investors would choose to incur trading costs so frequently and so simultaneously.

cannot explain the equity premium. They find that the consumption growth of stockholders does covary considerably more with stock returns than the consumption growth of nonstockholders. However, the covariance is *still* not large enough: it is only if they are highly risk averse that stockholders are marginally indifferent between stocks and bonds. There is still an equity premium puzzle when we look at the consumption of those involved in asset trade[27].

3. Summary

The Mehra-Prescott (1985) model assumes that asset trade is costless and that asset markets are complete. Casual observation suggests that both assumptions are unduly strong. Unfortunately, the predictions of the model for the equity premium and the riskfree rate are not greatly affected by allowing markets to be incomplete: dynamic self-insurance allows individuals to smooth idiosyncratic shocks without using explicit insurance contracts. It is true (by the law of supply and demand) that a frequently binding borrowing constraint forces the average riskfree rate down toward its sample value. The only way to simultaneously generate a sizeable equity premium, though, is to make trading in the stock market substantially more costly than in the bond market. Under this assumption, the premium on stocks represents compensation not for risk (as in the standard financial theory), but rather for bearing additional transactions costs. The sources of these additional costs remain unclear.

Despite its general lack of success in explaining the equity premium, this literature has made an important contribution to our understanding of trade in markets with frictions. Just as people in real life

---

[27]Beth Ingram (1990) examines the general equilibrium implications of a model in which half the agents use all possible information to form their portfolios, while the other half hold an optimal "nonchurning" portfolio: they are constrained not to change the split of assets between stocks and bonds in response to new information. She finds that the equity premium is no higher than if all agents were engaged in asset trade, but the riskfree rate is significantly lower. Presumably, the "rule of thumb" traders hold too many bonds in their portfolio because they want to be sure that they are insured against "runs" of bad income realizations. This drives the equilibrium interest rate downwards, but has little effect on the equity premium.

are able to find ways around government regulation of the economy, rational individuals in models are able to find ways around barriers to trade. It takes an enormous amount of sand in the gears to disrupt the ability of individuals to approximate an efficient allocation of resources.

There is, however, a continuing weakness in the "market frictions" literature. Usually, the incompleteness of markets and the costliness of trade in these models is motivated by adverse selection or moral hazard considerations. Yet, this motivation remains informal. From a theoretical point of view (and possibly an empirical one as well), the models will be stronger if they explicitly take into account the informational problems that lead to the trading frictions.

V. DISCUSSION AND CONCLUSIONS

This paper began by posing two puzzles: how can we explain why average stock returns are so much higher than bond returns in the United States data, and why has per capita consumption grown so quickly given that bond returns are so low? As it turns out, there are a variety of ways to generate more savings demand than in the Mehra-Prescott model and thereby explain the latter phenomenon.

However, the equity premium puzzle is much more challenging. Throughout this article, we have seen only two ways to explain the wedge in average returns between stocks and bonds. The first is that there is a large differential in the cost of trading between the stock and bond markets. To make this explanation compelling, it is important to ascertain the size of actual trading costs in these markets, and to provide an explanation of why those costs exist. Heaton and D. Lucas (1995b) and Aiyagari and Gertler (1991) discuss some early steps in this direction.

The second explanation is that, contrary to Mehra and Prescott's (1985) original parametric restrictions, individual investors have coefficients of relative risk aversion larger than ten (either with respect to their own consumption or with respect to per capita consumption)[28]. As I explained earlier, the problem with this explanation is that only a handful of economists believe that individuals are that risk averse. One way to support the "high risk aversion" view is to demonstrate that this apparently "strange" assumption about human behavior is consistent with data other than the average realization of the equity premium. Until now, little has been done along these lines, but Tallarini (1994)'s analysis of a "prototypical" real business cycle model using generalized expected utility preferences represents a promising first step.

Ten years ago, Mehra and Prescott (1985) wrote in their conclusion that the relatively low rate of return of Treasury bills, "is not the only example of some asset receiving a lower return than that implied by Arrow-Debreu general equilibrium theory. Currency, for example, is dominated by Treasury bills with positive nominal yields yet sizable amounts of currency are held." Thus, they regard the equity premium puzzle as being analogous to the so-called "rate of return dominance" puzzle that motivates much of modern monetary theory.

I believe that the past decade of work has made this analogy even more persuasive: I find that the suggested "solutions" to the equity premium puzzle closely resemble the approaches used by economists to generate a demand for money in the face of rate of return dominance. Thus, the "high risk aversion" story explains the puzzle by some peculiar aspect of individual preferences - which is exactly the way money-in-the-utility function models explain money demand. The "transactions costs" story explains that bonds are held despite their low rate of return because they are less costly to trade - which is exactly the way transactions costs models explain money demand.

Generally, both of these types of models of money demand are regarded as *ad hoc* "reduced forms". They fail to capture the essential *technological* forces (for example, lack of communication) that

[28]Of course, as mentioned above, to resolve the riskfree rate puzzle, we cannot only adopt a more flexible notion of what constitutes reasonable risk aversion; we must abandon the standard preferences (1) or impose borrowing constraints that bind frequently.

disrupt the process of fully centralized exchange and thereby generate a demand for money. This same complaint of superficiality can also be made of the two types of stories explaining the equity premium.

In one sense, the accuracy of the money demand analogy is depressing: monetary theorists are a long way from delivering a definitive model of money demand. On the other hand, the work of Robert Townsend (1987) and others in monetary theory is exciting because it shows so clearly what it will take to make true progress in explaining the size of the equity premium. Like fiat money, the equity premium appears to be a widespread and persistent phenomenon of market economies. The *universality* of the equity premium tells us that, like money, the equity premium must emerge from some primitive and elementary features of asset exchange that are probably best captured through extremely stark models.

With this in mind, we cannot hope to find a resolution to the equity premium puzzle by continuing in our current mode of patching the standard models of asset exchange with transactions costs here and risk aversion there. Instead, we must seek to identify what fundamental features of goods and asset markets lead to large risk-adjusted price differences between stocks and bonds. While I have no idea what these "fundamental features" are, it is my belief that any true resolution to the equity premium puzzle lies in finding them.

References

Abel, Andrew. "Asset Prices Under Habit Formation and Catching Up With the Joneses," Amer. Econ. Rev., May 1990, 80(2), pp. 38-42.

Abel, Andrew. "The Equity Premium Puzzle," Business Review: Federal Reserve Bank of Philadelphia, Sep./Oct. 1991, pp. 3-14.

Abel, Andrew. "Exact Solutions for Expected Rates of Return Under Markov Regime Switching: Implications for the Equity Premium Puzzle," J. Money, Credit, Banking, Part 1, Aug. 1994, 26(3), pp. 345-361.

Aiyagari, S. Rao. "Explaining Financial Market Facts: The Importance of Incomplete Markets and Transaction Costs," Fed. Res. Bank Minn. Rev., Winter 1993, 17(1), pp. 17-31.

__________. "Uninsured Idiosyncratic Risk and Aggregate Saving," Quart. J. Econ., Aug. 1994, 109(3), pp. 659-684.

Aiyagari, S. Rao and Gertler, Mark. "Asset Returns with Transaction Costs and Uninsured Individual Risk," J. Monet. Econ., June 1991, 27(3), pp. 311-331.

Allen, Edward R. "Evaluating Consumption-Based Models of Asset Pricing." Working Paper, The University of Chicago, 1990.

Atkeson, Andrew and Phelan, Christopher. "Reconsidering the Costs of Business Cycles with Incomplete Markets." Working Paper 4719, National Bureau of Economic Research, 1994.

Benninga, Simon and Protopapadakis, Aris. "Leverage, Time Preference and the 'Equity Premium Puzzle'," J. Monet. Econ., Jan. 1990, 25(1), pp. 49-58.

Blanchard, Olivier and Fischer, Stanley, eds. NBER Macroeconomics Annual 1992. Cambridge, MA: MIT Press, 1992.

Breeden, Douglas. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," J. Finan. Econ., Sep. 1978, 7(3), pp. 265-296.

Breeden, Douglas and Litzenberger, Robert. "State Contingent Prices Implicit in Option Prices," J. Bus., Oct. 1978, 51(4), pp. 621-651.

Breeden, Douglas; Gibbons, Michael and Litzenberger, Robert. "Empirical Tests of the Consumption-Oriented CAPM." J. Finance, June 1989, 44(2), pp. 231-262.

Burnside, Craig. "Hansen-Jagannathan Bounds as Classical Tests of Asset Pricing Models," J. Bus. Econ. Statist., Jan. 1994, 12(1), pp. 57-79.

Campbell, John and Cochrane, John. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." Working Paper, The University of Chicago, 1994.

Campbell, John. "Intertemporal Asset Pricing Without Consumption Data," Amer. Econ. Rev., June 1993, 83(2), pp. 487-512.

Carroll, Christopher. "The Buffer Stock Theory of Saving: Some Macroeconomic Evidence," Brookings Pap. Econ. Act., 1992, 2, pp. 61-135.

Cecchetti, Stephen and Mark, Nelson. "Evaluating Empirical Tests of Asset Pricing Models: Alternative Interpretations," Amer. Econ. Rev., May 1990, 80(2), pp. 48-51.

Cecchetti, Stephen; Lam, Pok-Sang and Mark, Nelson. "The Equity Premium and the Riskfree Rate: Matching the Moments," J. Monet. Econ., Feb. 1993, 31(1), pp. 21-45.

Cochrane, John and Hansen, Lars. "Asset Pricing Explorations for Macroeconomics," in Blanchard and Fischer 1992, pp. 115-182.

Constantinides, George. "Intertemporal Asset Pricing with Heterogeneous Consumers and Without Demand Aggregation," J. Bus., Apr. 1982, 55(2), pp. 253-267.

_____. "Habit Formation: A Resolution of the Equity Premium Puzzle," J. Polit. Econ., June 1990, 98(3), pp. 519-543.

Constantinides, George, and Duffie, Darrell. "Asset Pricing with Heterogeneous Consumers." Working Paper, The University of Chicago, 1995.

Deaton, Angus. Understanding Consumption. Oxford, UK: Clarendon Press, 1992.

den Haan, Wouter. "Heterogeneity, Aggregate Uncertainty, and the Short Term Interest Rate: A Case Study of Two Solution Techniques." Working Paper, University of California-San Diego, 1994.

Duesenberry, James. Income, Saving, and the Theory of Consumer Behavior. Cambridge, MA: Harvard University Press, 1952.

Epstein, Larry and Zin, Stanley. "Substitution, Risk aversion, and the Temporal Behavior of Consumption Growth and Asset Returns I: A Theoretical Framework, Econometrica, July 1989, 57(2), pp. 937-69.

_____. "First Order Risk Aversion and the Equity Premium Puzzle," J. Monet. Econ., Dec. 1990, 26(3), pp. 387-407.

_____. "Substitution, Risk Aversion, and the Temporal Behavior of Consumption Growth and Asset Returns II: An Empirical Analysis," J. Polit. Econ., Apr. 1991, 99(2), pp. 263-286.

Epstein, Larry and Wang, Tan. "Intertemporal Asset Pricing Under Knightian Uncertainty," Econometrica, Mar. 1994, 62(2), pp. 283-322.

Ferson, Wayne and Constantinides, George. "Habit Persistence and Durability in Aggregate Consumption: Empirical Tests," J. Finan. Econ., Oct. 1991, 29(2), pp. 199-240.

Gali, Jordi. "Keeping Up with the Joneses: Consumption Externalities, Portfolio Choice, and Asset Prices," J. Money, Credit, Banking, Feb. 1994, 26(1), pp. 1-8.

Giovannini, Alberto and Labadie, Pamela. "Asset Prices and Interest Rates in a Cash-in-Advance Economy," J. Polit. Econ., Dec. 1992, 99(6), pp. 1215-51.

Gregory, Allen and Smith, Gregor. "Calibration as Testing: Inference in Simulated Macroeconomic Models," J. Bus. Econ. Statist., July 1991, 9(3), pp. 297-303.

Grossman, Sanford and Shiller, Robert. "On the Determinants of Stock Market Variability," Amer. Econ. Rev., May 1981, 7(2), pp. 222-227.

Grossman, Sanford; Melino, Angelo and Shiller, Robert. "Estimating the Continuous-Time Consumption-Based Asset Pricing Model," J. Bus. Econ. Statist., July 1987, 5(3), pp. 315-327.

Haliassos, Michael and Bertaut, Carol. "Why Do So Few Hold Stocks? The View from Asset Pricing Theory." Working Paper, University of Maryland, 1991.

Hall, Robert. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," J. Polit. Econ., Dec. 1978, 86(6), pp. 971-987.

———. "Intertemporal Substitution and Consumption," J. Polit. Econ., Apr. 1988, 96(2), pp. 339-57.

Hamilton, James. "A New Approach to the Economic Analysis of Nonstationary Time Series and the

Business Cycle," Econometrica, Mar. 1989, 57(2), pp. 357-384.

Hansen, Lars; Heaton, John and Yaron, Amir. "Finite Sample Properties of Some Alternative GMM Estimators." Working Paper, The University of Chicago, 1994.

Hansen, Lars and Jagannathan, Ravi. "Implications of Security Market Data for Models of Dynamic Economies," J. Polit. Econ., Apr. 1991, 99(2), pp. 225-262.

Hansen, Lars; Sargent, Thomas and Tallarini, Thomas. "Pessimism, Neurosis and Feelings About Risk in General Equilibrium." Working Paper, The University of Chicago, 1994.

Hansen, Lars, and Singleton, Kenneth. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica, Sep. 1982, 50(5), pp. 1269-1286.

_____. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns, J. Polit. Econ., Apr. 1983, 91(2), pp. 249-265.

He, Hua and Modest, David. "Market Frictions and Consumption-Based Asset Pricing," J. Polit. Econ., Feb. 1995, 103(1), pp. 94-117.

Heaton, John. "An Empirical Investigation of Asset Pricing with Time Nonseparable Preferences," Econometrica, May 1995, 63(3), pp. 681-718.

Heaton, John and Lucas, Deborah. "Evaluating the Effects of Incomplete Markets on Risk Sharing and

Asset Pricing," J. Polit. Econ., 1995a (forthcoming).

_____. "The Importance of Investor Heterogeneity and Financial Market Imperfections for the Behavior of Asset Prices," Carnegie-Rochester Ser. Public Pol., 1995b (forthcoming).

Hubbard, Glenn; Skinner, Jonathan and Zeldes, Stephen. "The Importance of Precautionary Motives in Explaining Individual and Aggregate Saving," Carnegie-Rochester Ser. Public Pol., June 1994, 40, pp. 59-125.

Huggett, Mark. "The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies," J. Econ. Dynam. Control, Sep./Nov. 1993, 17(5/6), pp. 953-969.

Ingram, Beth. "Equilibrium Modeling of Asset Prices: Rationality Versus Rules of Thumb," J. Bus. Econ. Statist., Jan. 1990, 8(1), pp. 115-125.

Kandel, Shmuel and Stambaugh, Robert. "Expectations and Volatility of Consumption and Asset Returns," Rev. Finan. Stud., 1990, 3(2), pp. 207-232.

_____. "Asset Returns and Intertemporal Preferences," J. Monet. Econ., Feb. 1991, 27(1), pp. 39-71.

Kimball, Miles. "Precautionary Savings in the Small and in the Large," Econometrica, Jan. 1990, 58(1), pp. 53-73.

Kocherlakota, Narayana. "On the 'Discount' Factor in Growth Economies," J. Monet. Econ., Jan.

1990a, 25(1), pp. 43-48.

Kocherlakota, Narayana. "Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result," J. Finance, Mar. 1990b, 45(1), pp. 175-189.

_____. "On Tests of the Representative Agent Asset Pricing Model," J. Monet. Econ., Oct. 1990c, 26(2), pp. 285-304.

_____. "Testing the Consumption CAPM with Heavy-Tailed Pricing Errors." Working Paper, The University of Iowa, 1994.

Labadie, Pamela. "Stochastic Inflation and the Equity Premium," J. Monet. Econ., Sep. 1989, 24(3), pp. 277-298.

Lucas, Deborah. "Asset Pricing with Undiversifiable Risk and Short Sales Constraints: Deepening the Equity Premium Puzzle," J. Monet. Econ., Dec. 1994, 34(3), pp. 325-342.

Lucas, Robert E., Jr. "Asset Prices in an Exchange Economy," Econometrica, Sep. 1978, 46(5), pp. 1429-45.

_____. Models of Business Cycles. New York: Basil Blackwell, 1987.

Luttmer, Erzo. "Asset Pricing in Economies with Frictions." Working Paper, Northwestern University, 1994.

Mankiw, N. Gregory. "The Equity Premium and the Concentration of Aggregate Shocks," J. Finan. Econ., Oct. 1986, 17(2), pp. 211-219.

Mankiw, N. Gregory and Zeldes, Stephen. "The Consumption of Stockholders and Non-Stockholders." J. Finan. Econ., Mar. 1991, 29(1), pp. 97-112.

Marcet, Albert and Singleton, Kenneth. "Optimal Consumption-Savings Decisions and Equilibrium Asset Prices in a Model with Heterogeneous Agents Subject to Portfolio Constraints." Working Paper, Stanford University, 1991.

Mehra, Rajnish and Prescott, Edward. "The Equity Premium: A Puzzle," J. Monet. Econ., Mar. 1985, 15(2), pp. 145-161.

_____. "The Equity Premium: A Solution?" J. Monet. Econ., Jul. 1988, 22(1), pp. 133-136.

Nason, James. "The Equity Premium and Time-Varying Risk Behavior." Working Paper, Federal Reserve Board of Governors, 1988.

Neely, Christopher. "A Reconsideration of the Properties of the Generalized Method of Moments in Asset Pricing Models." Working Paper, Federal Reserve Bank of St. Louis, 1994.

Rietz, Thomas. "The Equity Risk Premium: A Solution," J. Monet. Econ., July 1988, 22(1), pp. 117-132.

Roy, Amlan. "Multicountry Comparisons of Consumption-Based Capital Asset Pricing Model: Germany, Japan, and U.S.A." Working Paper, University of London, 1994.

Rubinstein, Mark. "The Valuation of Uncertain Income Streams and the Pricing of Options," Bell J. Econ., Autumn 1976, 7(2), pp. 407-425.

Siegel, Jeremy. "The Real Rate of Interest from 1800-1990 in the United States and the United Kingdom," J. Monet. Econ., Apr. 1992, 29(2), 227-252.

Sotomayor, Marilda. "On Income Fluctuations and Capital Gains," J. Econ. Theo., Feb. 1984, 32(1), pp. 14-35.

Tallarini, Thomas. "Risk-Sensitive Real Business Cycles." Working Paper, The University of Chicago, 1994.

Telmer, Christopher. "Asset Pricing Puzzles and Incomplete Markets." J. Finance, Dec. 1993, 48(5), pp. 1803-1832.

Townsend, Robert. "Asset-Return Anomalies in a Monetary Economy." J. Econ. Theo., Apr. 1987, 41(2), pp. 219-247.

Weil, Philippe. "The Equity Premium Puzzle and the Riskfree Rate Puzzle," J. Monet. Econ., Nov. 1989, 24(2), pp. 401-421.

Weil, Philippe. "Nonexpected Utility in Macroeconomics," Quart. J. Econ., Feb. 1990, 105(1), pp. 27-42.

_____. "Equilibrium Asset Prices with Undiversifiable Labor Income Risk," J. Econ. Dynam. Control, July 1992, 16(3), pp. 769-790.

Zeldes, Stephen. "Consumption and Liquidity Constraints: An Empirical Investigation," J. Polit. Econ., Apr. 1989a, 97(2), pp. 305-346.

_____. "Optimal Consumption with Stochastic Income: Deviations from Certainty Equivalence," Quart. J. Econ., May 1989b, 104(2), pp. 275-298.

Appendix

Except for two exceptions, the following italicized description is taken directly from Mehra and Prescott (1985), pp. 147-48.

The data used in this study consists of five basic series for the period 1889-1978. The first four are identical to those used by Grossman and Shiller (1981) in their study. The series are individually described below.

(i) Series P: Annual average Standard & Poor's Composite Stock Price Index divided by the Consumption Deflator, a plot of which appears in Grossman and Shiller (1981, p. 225, fig. 1).

(ii) Series D: Real annual dividends for the Standard & Poor's series.

(iii) Series C: Kuznets-Kendrik-USNIA per capita real consumption on non-durables and services.

(iv) Series PC: Consumption deflator series, obtained by dividing real consumption in 1972 dollars on non-durables and serives by the nominal consumption on non-durables and services.

(v) Series RF: Nominal yield on relatively riskless short-term securities over the 1889-1978 period; the securities used were ninety-day government Treasury bills in the 1931-1978 period, Treasury Certificates for the 1920-1930 period and sixty-day to ninety-day Prime Commercial Paper prior to 1920 [the data was obtained from Homer (1963) and Ibbotson and Sinquefeld (1979)].

These series were used to generate the series actually utilized in this paper. Series P and D above were used to determine the average annual real return on the Standard & Poor's 500 Composite Index over the ninety-year period of study. The annual return for year t was computed as $(P_{t+1} + D_{t+1} - P_t)/P_t$. *The returns are plotted in fig. 2. Series C was used to determine the process on the growth rate of consumption over the same period... A plot of the percentage growth rate of real consumption appear in fig. 1. To determine the real return on a relatively riskless security we used the series RF and PC. For year t, this is calculated to be* $(1+RF_t)PC_t/PC_{t+1}$. *This series is plotted in fig. 3.*

This description differs from that of Mehra and Prescott (1985) in two ways. First, the figure numbers have been changed to match the numbering in this paper. Second, Mehra and Prescott (1985) used the formula $RF_t - (PC_{t+1} - PC_t)/PC_t$ for the real return to the relatively riskless security. This formula expresses the real rate as being the difference between the nominal rate and the inflation rate. Of course, their formula is only an approximation; I use the correct ratio formula instead.

This difference in the formulae creates only small differences in results. For example, they find that the average relatively riskless rate is 0.8%. I find it instead to be 1%. They find that the variance of the relatively riskless rate is 0.0032; I find it to be 0.00308.