

March 6, 2006

Understanding the Equity Risk Premium Puzzle

GEORGE M. CONSTANTINIDES*

ABSTRACT

The unconditional mean of the aggregate equity risk premium is almost six percent per year even after adjusting downwards the sample mean premium for unanticipated events in the latter part of the 20th century. In this essay I present theoretical and empirical research on three classes of generalizations of the standard neoclassical model and discuss their contribution towards a better understanding of the equity risk premium: preferences exhibiting habit persistence; borrowing constraints over the households' life cycle that limit capital market participation and concentrate the stock market risk on the saving middle-aged households; and the recognition that idiosyncratic income shocks are persistent, uninsurable and concentrated in economic recessions.

JEL classification: G12

*University of Chicago and NBER. I thank John Cochrane, Hanno Lustig, Roberto Violi and, particularly, Rajnish Mehra for their insightful comments and constructive criticism. I also thank Vito Gala and Lukasz Pomorski for their excellent research assistance and insightful comments. Naturally, I remain responsible for errors.

1. Introduction

The term “equity risk premium puzzle”, coined by Mehra and Prescott (1985), originally referred to the inability of standard neoclassical economic theory to reconcile the historically large realized premium of stock market return over the risk free rate with its low covariability with aggregate consumption growth. By now, the term encompasses a number of empirical regularities in the prices of capital assets that are at odds with the predictions of standard economic theory.

The neoclassical economic model parsimoniously links the returns of all assets to per capita consumption growth through the Euler equations of consumption.¹ According to the theory, the risk premia of financial assets are explained by their covariance with per capita consumption growth. However, per capita consumption growth covaries too little with the returns of most classes of financial assets and this creates a whole class of asset pricing puzzles: the aggregate equity return, the long-term bond return, and the returns of various subclasses of financial assets are too large, too variable, and too predictable. Attempts to leverage the low co-variability typically backfire, implying that the observed risk-free rate is too low and that its variance is too low. I discuss in some depth the aggregate equity puzzle because it exemplifies many of the problems that arise in attempting to explain the premium of any subclass of financial assets.

The covariance of per capita consumption growth with the aggregate equity return is *positive*. The rational model explains why the aggregate equity premium is positive. However, the covariance is typically one order of magnitude lower than what is needed to explain the premium. Thus, the equity premium is a *quantitative* puzzle.²

¹ See Breeden (1979), Lucas (1978), Merton (1973), and Rubinstein (1976)

² Ferson and Constantinides (1991), Grossman and Shiller (1981), Hansen and Jagannathan (1991), Hansen and Singleton (1982), and many others tested and rejected the Euler equations of consumption. Mehra and Prescott (1985) calibrated an economy to match the process of consumption growth. They demonstrated that the unconditional mean annual premium of the aggregate equity return over the risk-free rate is, at most, 0.35 percent. This is too low, no matter how one estimates the unconditional mean equity premium. Weil (1989) stressed that the puzzle is a dual puzzle of the observed too high equity return and too low risk-free rate.

The equity premium puzzle is *robust*. One may address the problem by testing the Euler equations of consumption or by calibrating the economy. Either way, it is a puzzle. In calibrating an exchange economy, the model cannot generate the first and second unconditional moments of the equity returns. In testing and rejecting the Euler equations of consumption, one abstracts from the market clearing conditions. Variations in the assumptions on the supply side of the economy do not resolve the puzzle.

The challenge is a *dual puzzle* of the equity premium that is too high and the risk-free rate that is too low relative to the predictions of the model. In calibrating an economy, the strategy of increasing the risk aversion coefficient in order to lever the effect of the problematic low covariance of consumption growth with equity returns increases the predicted risk-free rate and aggravates the risk-free-rate puzzle. In testing the Euler equations of consumption, the rejections are strongest when the risk free rate is included in the set of test assets.

Several generalizations of essential features of the model have been proposed to mitigate its poor performance. They include alternative assumptions on preferences;³ rare but disastrous market-wide events;⁴ distorted beliefs and learning;⁵ incomplete markets;⁶ market imperfections;⁷ and liquidity risk.⁸ They also include a better understanding of data problems such as limited participation of consumers in the stock

³ For example, Abel (1990), Bansal and Yaron (2004), Benartzi and Thaler (1995), Boldrin, Christiano, and Fisher (2001), Braun, Constantinides and Ferson (1993), Cagetti (2002), Campbell and Cochrane (1999), Constantinides (1990), Epstein and Zin (1991), and Ferson and Constantinides (1991).

⁴ For example, Barro (2005), Danthine and Donaldson (1999), Mehra and Prescott (1988), and Rietz (1988).

⁵ Cecchetti, Lam and Mark (2000), Danthine and Donaldson (1999), Guidolin (2004), and Weitzman (2005).

⁶ For example, Bewley (1982), Brav, Constantinides, and Geczy (2002), Constantinides and Duffie (1996), De Santis (2004), Heaton and Lucas (1996), Jacobs and Wang (2004), Krebs (2002), Mankiw (1986), Mehra and Prescott (1985), and Storesletten, Telmer, and Yaron (2001, 2004).

⁷ For example, Aiyagari and Gertler (1991), Bansal and Coleman (1996), Constantinides, Donaldson, and Mehra (2002), Daniel and Marshall (1997), Danthine, Donaldson, and Mehra (1992), Guvenen (2005), He and Modest (1995), Heaton and Lucas (1996), and Parker and Julliard (2005).

⁸ See, Alvarez and Jermann (2001) and Lustig (2004).

market;⁹ temporal aggregation;¹⁰ and the survival bias of the United States capital market.¹¹

The low covariance of the growth rate of aggregate consumption with equity returns is a major stumbling block in explaining the mean aggregate equity premium and the cross-section of asset returns, in the context of a representative-consumer economy with time separable preferences. Mankiw and Shapiro (1986) found that the market beta often explains asset returns better than the consumption beta does. Over the years, a number of different economic models have been proposed that effectively increase the covariance of equity returns with the growth rate of aggregate consumption, by proxying the growth rate of aggregate consumption with the stock market return in the Euler equations of consumption.¹²

Many of these generalizations contribute in part toward our better understanding of the economic mechanism that determines the pricing of assets. I refer the reader to the essays in this Handbook; the textbooks by Campbell, Lo, and McKinlay (1997) and Cochrane (2001); and the articles by Campbell (2001, 2003), Cochrane (1997), Cochrane and Hansen (1992), Constantinides (2002), Kocherlakota (1996), and Mehra and Prescott (2003).

My objective in this essay is to provide a better understanding of these issues when I generalize, one at a time, three of the key assumptions of the standard theory. First, I relax the assumption that preferences are time separable and introduce preferences

⁹ For example, Attanasio, Banks and Tanner (2002), Brav, Constantinides, and Geczy (2002), Brav and Geczy (1995), Mankiw and Zeldes (1991), and Vissing-Jorgensen (2002).

¹⁰ See, Gabaix and Laibson (2001), Heaton (1995), and Lynch (1996).

¹¹ See Brown, Goetzmann, and Ross (1995). However, Jorion and Goetzmann (1999, Table 6) found that the average real capital gain rate of a United States equities index exceeds the average rate of a global equities index that includes both markets that have and have not survived by merely one percent per year.

¹² Friend and Blume (1975) explained the mean equity premium with low RRA coefficient by assuming a single-period economy in which the end-of-period consumption inevitably equals the end-of-period wealth. In the Epstein and Zin (1991) model, even though the preferences are defined over consumption alone, the stock market return enters directly in the Euler equations of consumption. Bakshi and Chen (1996) introduced a set of preferences defined over consumption and wealth—the spirit of capitalism—that also have the effect of introducing the stock market return in the Euler equations of consumption.

that exhibit habit persistence. Second, I recognize that participation in the stock market by households is limited and that some aggregate consumption measures reflect the consumption of both participating and non participating households in the stock market. Lastly, I relax the assumption that the market is complete and address the implications of incomplete consumption insurance.

I begin with the standard assumptions of neoclassical economic theory, as adapted in finance. There are I consumers $i, i = 1, \dots, I$, each consuming c_t^i units of the consumption good in period t and having von Neumann-Morgenstern preferences

$$E_0 [U_i (c_0^i, \dots, c_t^i, c_{t+1}^i, \dots)]. \quad (1.1)$$

There are J capital assets, $j = 1, \dots, J$, traded by the consumers in perfect markets. Prices, dividends and income are denominated in units of the consumption good. The *ex-dividend* price of the j^{th} asset at time t is $P_{j,t}$, its dividend is $D_{j,t}$, and its return between dates t and $t+1$ is $R_{j,t+1} = (P_{j,t+1} + D_{j,t+1}) / P_{j,t}$. The i^{th} consumer's budget constraint at time t is

$$c_t^i + y_t^i + \sum_{j=i}^J \theta_{j,t}^i P_{j,t} \leq \sum_{j=i}^J \theta_{j,t-1}^i (P_{j,t} + D_{j,t}) \quad (1.2)$$

where y_t^i is non-dividend income at time t and $\theta_{j,t}^i$ is the fraction of the j^{th} asset held by the i^{th} consumer after trading at time t .

The i^{th} consumer maximizes expected utility by choosing a consumption and investment plan that satisfies the budget constraints. By definition, if (c_t^i, c_{t+1}^i) is the optimal consumption plan at times t and $t+1$, then the feasible consumption plan $(c_t^i - \delta, c_{t+1}^i + \delta R_{j,t+1})$ maximizes expected utility with respect to δ at $\delta = 0$:

$$\left. \frac{\partial E_0 [U_i(c_0^i, \dots, c_t^i - \delta, c_{t+1}^i + \delta R_{j,t+1}, \dots)]}{\partial \delta} \right|_{\delta=0} = 0. \quad (1.3)$$

This variational argument leads to the following Euler equation of consumption between times t and $t+1$:

$$E_t \left[-\frac{\partial U_i}{\partial c_t^i} + \frac{\partial U_i}{\partial c_{t+1}^i} R_{j,t+1} \right] = 0. \quad (1.4)$$

In the standard theory, preferences are typically specialized to be time separable as

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t u_i(c_t^i) \right] \quad (1.5)$$

with constant subjective discount factor β . There is no empirical justification for this specialization of preferences. In Section 2, I relax this assumption, introduce preferences that exhibit *habit persistence* and explore its ramifications in understanding the observed equity premium.

In any case, with time separable preferences as in equation (1.5), the Euler equation of consumption simplifies into the following equation:

$$E_t \left[\beta \frac{\partial u_i(c_{t+1}^i) / \partial c_{t+1}^i}{\partial u_i(c_t^i) / \partial c_t^i} R_{j,t+1} \right] = 1. \quad (1.6)$$

In the standard theory, the model is typically further specialized by assuming that the market is complete. There is little, if any, empirical justification for this assumption either. In Section 4, I relax this assumption, considering instead an *incomplete market*, and explore its implications in addressing the observed equity premium.

In any case, under the assumption of market completeness, the equilibrium in this heterogeneous-consumer economy is isomorphic in its pricing implications to the

equilibrium in a representative-consumer economy.¹³ For our purposes, market completeness implies that there exists a period utility function $u(C_t)$ of per capita consumption $C_t \equiv I^{-1} \sum_{i=1}^I c_t^i$, such that we may replace the I consumer-specific Euler equations (1.6) with the following Euler equation of consumption of the representative consumer:

$$E_t \left[\beta \frac{\partial u(C_{t+1}) / \partial C_{t+1}}{\partial u(C_t) / \partial C_t} R_{j,t+1} \right] = 1. \quad (1.7)$$

It is often assumed that the period utility function is of the power form, $u(C_t) = (1-A)^{-1} (C_t)^{1-A}$, with constant relative risk aversion (RRA) coefficient A . With power utility, the Euler equation (1.7) further simplifies into the following:

$$E_t \left[\beta \left(\frac{C_{t+1}}{C_t} \right)^{-A} R_{j,t+1} \right] = 1. \quad (1.8)$$

An advantage of the assumption that the period utility function is of the power form is that per capita consumption appears in the Euler equation only as per capita consumption *growth*, C_{t+1}/C_t , which may be stationary even if the consumption *level* is non-stationary.

In empirical work and calibration, per capita consumption growth in the Euler equation (1.8) is often taken to be the aggregate consumption growth, where the aggregate consumption is taken from the National Income and Product Accounts (NIPA). However, the aggregate consumption reported by NIPA is the sum total over all households, irrespective of whether these households are marginal in the equities and bond markets or not. In Section 3, I relax this assumption, recognizing instead the *limited stock market participation* by households and explore its ramifications in addressing the observed equity premium. I pursue these issues in the context of a lifecycle model in

¹³ See Section 4 for an elaboration of this.

which young consumers are constrained from borrowing against their future labor income to invest in the stock market.

Leaving aside the issue of limited capital market participation, I condition down the Euler equation (1.8) into unconditional form and linearize it under the assumption that consumption growth and asset returns are bivariate lognormally distributed:

$$\begin{aligned} \ln E[R_{i,t+1}] &\equiv E[\ln R_{i,t+1}] + \frac{1}{2} \text{var}(\ln R_{i,t+1}) \\ &= -\ln \beta + A \times E\left[\ln \frac{C_{t+1}}{C_t}\right] - \frac{1}{2} A^2 \times \text{var}\left(\ln \frac{C_{t+1}}{C_t}\right) + A \times \text{cov}\left(\ln R_{i,t+1}, \ln \frac{C_{t+1}}{C_t}\right) \end{aligned} \quad (1.9)$$

The term $\frac{1}{2} A^2 \times \text{var}\left(\ln \frac{C_{t+1}}{C_t}\right)$ reflects the decrease in the expected rate of return due to the precautionary demand for savings. The term $A \times \text{cov}\left(\ln R_{i,t+1}, \ln \frac{C_{t+1}}{C_t}\right)$ reflects the “consumption-beta” risk premium.

I take the difference of the linearized Euler equations for the market return and for the risk free rate and find that the unconditional mean equity premium is equal to the product of the RRA coefficient of the representative consumer and the covariance of the market return with per capita consumption growth:¹⁴

$$\ln E[R_{M,t+1}] - \ln E[R_{F,t+1}] = A \times \text{cov}\left(\ln R_{M,t+1}, \ln \frac{C_{t+1}}{C_t}\right). \quad (1.10)$$

I also find that the risk free rate is equal to the sum of the subjective discount rate $(-\ln \beta)$ and the product of the RRA coefficient and the mean growth rate in

¹⁴ I use the property that the risk free rate $R_{F,t+1}$ is known at time t . The *conditional* covariance $\text{cov}_t(\ln R_{F,t+1}, \ln(C_{t+1}/C_t))$ is zero and, therefore, the *unconditional* covariance $\text{cov}(\ln R_{F,t+1}, \ln(C_{t+1}/C_t))$ is approximately zero.

consumption, less one-half the product of the squared RRA coefficient and the variance of consumption growth:

$$\ln E[R_{F,t+1}] = -\ln \beta + A \times E\left[\ln \frac{C_{t+1}}{C_t}\right] - \frac{1}{2}A^2 \times \text{var}\left(\ln \frac{C_{t+1}}{C_t}\right). \quad (1.11)$$

In this context, the *elasticity of substitution in consumption* is the increase in expected consumption growth for a unit increase in the risk free rate, $\psi \equiv \partial \ln E[C_{t+1}/C_t] / \partial \ln E[R_{F,t+1}]$.¹⁵ With time-separable preferences and constant RRA coefficient, equation (1.11) implies that the product of the RRA coefficient and the elasticity equals one, $RRA \times \psi = 1$. In terms of the elasticity, equation (1.11) may be written as

$$\ln E[R_{F,t+1}] = -\ln \beta + \frac{1}{\psi} \times E\left[\ln \frac{C_{t+1}}{C_t}\right] - \frac{1}{2\psi^2} \times \text{var}\left(\ln \frac{C_{t+1}}{C_t}\right). \quad (1.12)$$

With time-separable preferences and constant RRA coefficient, equations (1.11) and (1.12) are interchangeable.¹⁶

Empirically, the covariance of per capita consumption growth with the market return is low and cannot generate the mean equity premium, as in equation (1.10), with a

¹⁵ This is a specialization of the general definition of the elasticity of substitution in consumption,

$$\psi \equiv \partial \ln[C_{t+j}/C_t] / \partial \ln(u'_{t+j}/u'_t).$$

¹⁶ With time-separable preferences that allow variable RRA coefficient, $u(C_t) = (1-A)^{-1}(C_t - K)^{1-A}$,

equation (1.11) is replaced by $\ln E[R_{F,t+1}] = -\ln \beta + A \times E\left[\ln \frac{(C_{t+1} - K)}{(C_t - K)}\right] - \frac{1}{2}A^2 \times \text{var}\left(\ln \frac{(C_{t+1} - K)}{(C_t - K)}\right)$. It

is still the case that $A \times \psi \approx 1$. However A no longer equals the RRA coefficient and the product $RRA \times \psi$ need not equal one. Also, with time-non-separable preferences, such as habit-forming preferences, and state-non-separable preferences as in Epstein-Zin (1991), the product $RRA \times \psi$ need not equal one. With Epstein-Zin (1991) preferences, the risk free rate is given by equation (1.12) in terms of the elasticity and not by equation (1.11) in terms of the RRA coefficient.

reasonable value of the RRA coefficient. If one assumes a sufficiently high value of the RRA coefficient in order to generate a realistic mean premium, the Euler equation of consumption with respect to the risk free rate, equation (1.11), now generates a risk free rate which is much higher than the historically observed risk free rate. Thus the puzzle is a dual puzzle that the observed mean equity premium is too high and the risk free rate is too low.¹⁷ These are but two of the implications of the Euler equations (1.7) that are at odds with the data either in empirical work or calibration.

In the following sections, I sequentially examine preferences that exhibit habit persistence, recognize the limited stock market participation by households, and relax the assumption of market completeness and explore their ramifications in better understanding the equity premium, the risk free rate and the Euler equations of consumption.

¹⁷ Note that the risk free rate is decreasing in the RRA coefficient for sufficiently high value of this coefficient. However, it takes an absurdly high value of the RRA coefficient to lower the risk free rate in this way.

2. Habit persistence

I explore the contribution of habit persistence in resolving the equity premium puzzle and related anomalies. Habit persistence has a long tradition in economic theory, dating back to Marshall (1920) and Duesenberry (1949). It is the property of preferences that an increase in consumption increases the marginal utility of consumption at adjacent dates relative to the marginal utility of consumption at distant ones.

I model *internal habit*, x_t , as an exponentially weighted sum of the representative consumer's own past consumption, c_t , as

$$x_t = e^{-at} x_0 + b \int_0^t e^{a(s-t)} c_s ds. \quad (2.1)$$

where the parameter a is the persistence of habit.

Duesenberry (1949) introduced a model of *external habit persistence* in which the consumer does not take into account the effect of her own current consumption on her future preferences but rather takes into account the effect of other consumers' consumption on her future preferences. I model *external habit*, X_t , as an exponentially weighted sum of per capita past consumption, C_t , of *identical* consumers as

$$X_t = e^{-a't} X_0 + b' \int_0^t e^{a'(s-t)} C_s ds. \quad (2.2)$$

In models with external habit, one typically assumes that there are a large number of identical consumers so that a consumer chooses her optimal consumption taking the path of per capita consumption as exogenous. Naturally, in equilibrium, $c_t = C_t$.

I switch from a discrete-time to a continuous-time framework and specialize the von Neumann-Morgenstern preferences stated in equation (1.1) as follows:

$$\int_0^\infty e^{-\rho t} (1-A)^{-1} (f(c_t, x_t, X_t))^{1-A} dt \quad (2.3)$$

where $\rho \equiv -\ln \beta$ is the subjective discount rate. More generally, I may define $f(\cdot)$ to be a function of internal and external habits at several different frequencies. The concavity parameter A has the interpretation as the RRA coefficient in the time separable model, $f(c_t, x_t, X_t) = c_t$, but not generally so in the habit model even if $f(\cdot)$ is linear in its arguments.

Building on earlier work by Ryder and Heal (1973) and Sundaresan (1989), I demonstrated in Constantinides (1990) that internal habit persistence can, in principle, reconcile the high mean equity premium with the low variance of consumption growth and with the low covariance of consumption growth with equity returns. In this model, habit is *internal* and *linear*:

$$f(c_t, x_t, X_t) = c_t - x_t. \quad (2.4)$$

From equations (2.3) and (2.4), the period utility is $u(c_t, x_t) = (1 - A)^{-1} (c_t - x_t)^{1-A}$.

I embed these preferences in a representative-consumer production economy and discuss their implications. There exists only one production good that is also the consumption good. This good may be consumed or invested without adjustment costs in two constant-returns-to-scale technologies. The technologies have constant returns to scale and rates of return over the period $[t, t + dt]$ equal to $r dt$ and $\mu dt + \sigma dw(t)$, respectively, where r, μ and σ are constants and $w(t)$ is a standard Brownian motion in R^1 . I do not explicitly model the labor input to production because my focus is on the Euler equations of consumption and their implications on the equity premium puzzle.

Since this is a representative-consumer economy, I interpret the optimal consumption and investment paths as the equilibrium paths and the shadow prices of assets as the equilibrium prices. In calibrating the model, I choose parameter values such that both technologies are always active. Then the rate of return of the risk free technology determines the risk free rate as r ; and the parameters of the risky technology determine the mean and standard deviation of the return on the stock market as μ and

σ , respectively. I do not model the consumers' labor income. Therefore, consumption equals dividends plus interest, net of new investment in production. One way to rationalize the assumption that consumers do not receive labor income is to view the capital invested in production as including human capital, the capitalized value of future labor income. Then the return on the risky technology is interpreted as the return on both the stock market and human capital.

There is a single state variable in this economy, the habit-to-consumption ratio, x_t/c_t . In Constantinides (1990), I solved this model. The wealth, W_t , and consumption dynamics are

$$\frac{W_t - \frac{x_t}{r+a-b}}{W_0 - \frac{x_0}{r+a-b}} = \frac{c_t - x_t}{c_0 - x_0} = \exp \left\{ \left(n - \frac{m^2 \sigma^2}{2} \right) t + m \sigma w(t) \right\} \quad (2.5)$$

where

$$n \equiv \frac{r - \rho}{A} + \frac{(\mu - r)^2 (1 + A)}{2A^2 \sigma^2} \quad (2.6)$$

and

$$m = \frac{\mu - r}{A \sigma^2}. \quad (2.7)$$

Consumption growth is

$$\frac{dc_t}{c_t} = \left\{ n + b - (n + a) \left(\frac{x_t}{c_t} \right) \right\} dt + \left\{ 1 - \frac{x_t}{c_t} \right\} m \sigma dw(t) \quad (2.8)$$

and the derived utility of wealth is

$$V(W_t, x_t, t) = \theta \times \left(W_t - \frac{x_t}{r+a-b} \right)^{1-A} \quad (2.9)$$

where θ is a constant.

The Arrow-Pratt relative risk aversion coefficient is defined over atemporal gambles in terms of the derived utility of wealth as

$$\begin{aligned}
 RRA &= -\frac{WV_{WW}}{V_W} = \frac{A}{1 - \frac{x_t/W_t}{r+a-b}} \\
 &= A \left\{ 1 + \frac{\rho - (1-A)r - \frac{(1-A)(\mu-r)^2}{2A\sigma^2}}{A(r+a)} \times \frac{x_t}{c_t - x_t} \right\}. \tag{2.10}
 \end{aligned}$$

The RRA coefficient is not given by the expression $-c_t u_{cc} / u_c = A / (1 - x_t / c_t)$.¹⁸ In the calibration below, the RRA coefficient is very close to the parameter A , even though the expression $A / (1 - x_t / c_t)$ is much higher than A .

In calibrating the model, I choose the technology parameters r, μ and σ such that the annual risk free rate is 1% and the annual market return has mean 7% and standard deviation 16.5%, implying a mean annual equity premium of 6%. The goal is to choose the remaining parameters ρ, A, a , and b so that annual consumption growth rate has mean 1.8% and standard deviation 3.6% while the RRA coefficient is kept at a

¹⁸ If the utility function is time-separable with period utility $(1-A)^{-1} c_t^{1-A}$, then $V(W_t, t) = h W_t^{1-A}$, where h is a function of time and the state variables. In this case, the RRA coefficient exactly equals $-c u_{cc} / u_c$ because $-W_t V_{WW} / V_W = A$ and $-c u_{cc} / u_c = A$. If the utility function is time-separable with period utility $(1-A)^{-1} (c_t - \hat{c})^{1-A}$ and \hat{c} is a constant, then $V(W_t, t) = h (W_t - \hat{c}/r)^{1-A}$, if the risk free rate r is constant; as before, h is a function of time and the state variables. In this case, the RRA coefficient is $-W_t V_{WW} / V_W = A / (1 - \hat{c}/r W_t)$ and is approximately equal to $-c u_{cc} / u_c = A / (1 - \hat{c}/c_t)$, only if $(1 - \hat{c}/r W_t) / (1 - \hat{c}/c_t) \approx 1$.

plausibly low value.¹⁹ The order of magnitude of the subjective discount factor is not a controversial parameter, so I set $\rho = 0.037$ per year.

Before I discuss the calibration of the habit model, I demonstrate that the time separable model fails to achieve the calibration goals. To see this, I set $b = 0$. If I want the model to imply that 100% (or 50%) of the wealth is invested in the stock market, I must set the RRA coefficient at 2.2 (or, 4.4). Then the model implies that the annual consumption growth has mean 3% (or, 1%) and standard deviation 16.5% (or, 8.2%). In order to generate the moments of the market return and the risk free rate, the model requires much higher variability of consumption growth and covariability with the market return than what is observed in the data.

With internal habit, several parameter combinations attain the calibration goal, as displayed in Constantinides (1990, Table 1). For example, with parameters $A = 2.2$, $a = 0.5$, and $b = 0.405$, the RRA coefficient has unconditional mean 2.88, and the annual consumption growth has unconditional mean 1.8% and standard deviation 3.6%. On average, 79% of the wealth is invested in the stock market. Furthermore, consumption growth is substantially less variable than the growth in wealth. For example, with the above parameter values, the standard deviation of the instantaneous consumption growth is one-fifth of that of the growth in wealth.

An intuition behind the internal habit model is that the RRA coefficient in equation (2.10) is increasing in the habit-to-consumption ratio and is, therefore, countercyclical. To see this, I first discuss a common property of economies with time-separable preferences and internal habit preferences. In an economic recession, equity returns are low and this event typically occurs when consumption is low and, therefore, the marginal utility of consumption is high. This effect alone is insufficiently strong to generate the order of magnitude of the observed premium, if the RRA coefficient is constant, as in the standard model with time-separable preferences. In the habit model, the decrease in consumption increases the habit-to-consumption ratio, increases the RRA aversion coefficient, and raises the marginal utility of consumption much higher than in a

¹⁹ This exercise is the “dual” of that in Mehra and Prescott (1985). In that paper, the authors started with a consumption process with a mean growth rate of 1.8% and standard deviation of 3.6% and sought reasonable parameter values to rationalize the observed equity premium.

model with time separable preferences. It turns out that the combination of the two effects is sufficiently strong to generate the order of magnitude of the observed premium without an unreasonably high RRA coefficient. The same intuition drives the non-linear external habit model of Campbell and Cochrane (1999) which I discuss later on.

Another intuition behind the internal habit model may be gleaned by studying the relationship between the RRA coefficient and the elasticity of substitution in consumption. Recall our earlier discussion of the equity premium puzzle and the risk free rate puzzle through equations (1.8) and (1.10). In the time-separable model with constant RRA coefficient, the product of the RRA coefficient and the elasticity, ψ , equals one. A high value of the RRA coefficient is needed to generate the observed equity premium. However, the high value of the RRA coefficient implies a low value for the elasticity and an unrealistically large value of the risk free rate. By contrast, in the habit model, the product $RRA \times \psi$ can be substantially lower than one, as shown in Constantinides (1990, Table 1). For the particular choice of parameter values that I discussed earlier on, the product $RRA \times \psi$ is about 0.26.

Epstein and Zin (1991) put forth a model of preferences outside the class of von Neumann-Morgenstern preferences. Their model has the advantage that the RRA coefficient and the elasticity of substitution in consumption are neatly represented as two distinct parameters. Their model disentangles these two important economic concepts, with the former driving the equity premium and the latter driving the risk free rate. However, the lesson derived from our earlier discussion of the habit model is that departure from von Neumann-Morgenstern preferences is not a necessary condition for disentangling these two important economic concepts. Even within the class of von Neumann-Morgenstern preferences, departure from time separability in the form of habit persistence suffices to disentangle the RRA coefficient and the elasticity of substitution in consumption.

A limitation of the internal habit model is that it implies autocorrelations much higher than the observed ones at monthly, quarterly and annual frequencies. The model also implies standard deviations somewhat higher than the observed ones. Whereas the model-implied *instantaneous* consumption growth is serially uncorrelated, the k -period consumption growth is not. I define the k -period consumption growth rate as

$\ln \left\{ \left(\int_{t+k}^{t+2k} c_s ds \right) / \left(\int_t^{t+k} c_s ds \right) \right\}$ and numerically calculate its standard deviation and first-order autocorrelation for $k = 1, 3,$ and 12 months and parameter values $A = 2.2, a = 0.5,$ and $b = 0.405$. The results are reported in the table below.

	1 month	3 months	12 months
Model-implied mean std dev	0.013	0.026	0.070
Ferson-Constantinides estimated std dev	0.008	0.022	0.031
Model-implied mean autocorrelation	0.383	0.531	0.693
Ferson-Constantinides estimated autocorr.	- 0.36	- 0.31	0.41

In the same table, I also present the monthly estimates for the period 1959-1986, the seasonally adjusted quarterly estimates for the period 1948-1986 and the annual estimates for the period 1930-1986, as reported in Ferson and Constantinides (1991). It should be pointed out that the sample autocorrelation of the consumption growth rate varies widely over different samples. For example, the Ferson and Constantinides (1991) estimate of 0.41 autocorrelation of annual consumption growth for the period 1930-1986 contrasts with the Mehra and Prescott (1985) estimate of -0.14 for the period 1889-1978. Despite the wide range of these estimates, the internal habit model implies autocorrelations much higher than the observed ones.

Campbell and Cochrane (1999) argued that the annual endowment growth process (and, therefore, the equilibrium consumption growth process) should be close to i.i.d.. However, an exchange economy with internal habit and i.i.d. annual endowment growth process yields a now endogenous risk free rate that is excessively variable. This observation partly motivates the non-linear external habit model of Campbell and Cochrane (1999) that generates a constant endogenous risk free rate when the annual endowment growth process as i.i.d. I discuss their model later on.

The calibration results of the internal habit model are reinforced by empirical evidence. Braun, Constantinides and Ferson (1993) and Ferson and Constantinides (1991) tested the Euler equations of consumption in the special case of the linear internal habit model in which the habit depends only on the *first* lag of own consumption. They reported that the habit model performs better than the time-separable model and that the habit persistence parameter is economically and statistically significant.

In what sense does the linear internal habit model fit the data better than the time separable model? Ferson and Constantinides (1991, Table 11) added a parameter L to the risk free rate and re-estimated the parameters, including the parameter L , from the Euler equations. For the 5-asset system of Euler equations (Treasury bills, government bonds, and equity size deciles 1, 5 and 10), the estimated value of L is small and insignificantly different from zero. However, for the 2-asset system of Euler equations (Treasury bills and equity size decile 10), the estimated value of L is about 6% annual and significantly different from zero. More to the point, the estimated value of L is the same whether Ferson and Constantinides (1991) allow for habit in the Euler equations or set to zero the habit parameters. The fit to the data is better with habit than without. However, the improvement in the fit is not in better accommodating the mean equity premium. In both the time-separable and the linear internal habit model, they estimated an unexplained premium of about 6% annual. The interpretation of this little-noticed result is that habit persistence improves the fit of the model largely through its effect on moments other than the mean equity premium.

Cecchetti, Lam, and Mark (1994), Cochrane and Hansen (1992), Gallant, Hansen and Jagannathan (1991), Hansen and Tauchen (1990), and Heaton (1995) found that the linear internal habit model brings the moments of the marginal rate of substitution closer to the feasible Hansen and Jagannathan (1991) region than the time separable model does.

In exploring the differences between internal and external habit models, it is convenient to switch back to a discrete-time setup and state the preferences as

$$\sum_{t=0}^{\infty} \beta^t u(c_t, x_t, X_t). \quad (2.11)$$

I model *internal habit*, x_t , as an exponentially weighted sum of the representative consumer's own past consumption, c_t , over S lags as

$$x_t = b \sum_{s=1}^S \phi^s c_{t-s}. \quad (2.12)$$

I model *external habit*, X_t , as an exponentially weighted sum of per capita past consumption, C_t , of *identical* consumers over S lags as

$$X_t = b \sum_{s=1}^S \varphi^s C_{t-s}. \quad (2.13)$$

With external habit, $u(c_t, x_t, X_t) = u(c_t, X_t)$, the marginal utility of consumption at t is

$$MU_t^{EXT} = \beta^t \frac{\partial u(c_t, X_t)}{\partial c_t}. \quad (2.14)$$

By contrast, with internal habit, $u(c_t, x_t, X_t) = u(c_t, x_t)$, the marginal utility of consumption at time t is

$$MU_t^{INT} = \beta^t \frac{\partial u(c_t, x_t)}{\partial c_t} + E_t \left[\sum_{s=1}^S \beta^{t+s} \frac{\partial u(c_{t+s}, x_{t+s})}{\partial x_{t+s}} \frac{\partial x_{t+s}}{\partial c_t} \right]. \quad (2.15)$$

The Euler equation of consumption for an asset with return R_{t+1} is

$$MU_t = E_t [MU_{t+1} \times R_{t+1}]. \quad (2.16)$$

With external habit, the contemporaneous consumption growth c_{t+1}/c_t and its S lags appear in the Euler equations of consumption. By contrast, with internal habit, the S forward values of consumption growth, in addition to the contemporaneous consumption growth c_{t+1}/c_t and its S lags appear in the Euler equations of consumption. Thus, external and internal habit models are empirically distinguishable.²⁰

²⁰ Hansen and Sargent (undated) and Cochrane (2001) pointed out that the pricing implications of internal and external habit models are identical, in the special case when *habit is linear* and *the risk free rate is constant*. In the external habit model with period utility $u(c_t, X_t) = u(c_t - X_t)$, the marginal utility is

Turning to the empirical evidence, Bansal, Dittmar and Lundblad (2004), Bansal and Yaron (2004), Hansen, Heaton and Li (2005), Malloy, Moskowitz and Vissing-Jorgensen (2005) and Parker and Julliard (2005) reported that forward-looking consumption growth explains the cross-section of stock returns better than contemporaneous consumption growth does. These results may be interpreted as evidence in support of the internal habit models and against the external habit models. However, factors other than internal habit may be driving these results. They include measurement error, adjustment costs, nonseparability of marginal utility of consumption from factors such as labor supply or housing stock, Epstein-Zin (1991) preferences and behavioral factors. Ferson and Harvey (1992) tested the linear external habit model and reported positive results. Chen and Ludvigson (2005) treated the functional form of nonlinear habit as an unknown and estimated it along with the other model parameters from the data in both the internal habit and external habit versions of the model. They concluded that habit is nonlinear, is internal and explains the cross-section of stock returns better than most extant factor models do.

Two specific non-linear habit models have received attention in the literature. The first, by Abel (1990) is a model of non-linear (multiplicative) internal and external habit, $f(c_t, x_t, X_t) = (c_{t-1}^D C_{t-1}^{1-D})^{-1} c_t$, where D is a parameter. With external habit, $D = 0$, Abel (1990) found that the unconditional expected equity premium and risk free rate are close to their sample values; however, the conditional expected values are too variable. With internal habit, $D = 1$, his results are inconclusive and oversensitive to the parameter values. This model is discussed in depth in Mehra's essay in this handbook.

Campbell and Cochrane (1999) modeled an exchange economy with i.i.d. annual endowment growth process and preferences in which the habit is external and is a non-linear function of past changes in consumption rather than an exponentially-weighted

$MU_t^{EXT} = \beta' u'(c_t - X_t)$. If the Euler equation for the risk free asset holds with external habit, then $u'(c_t - X_t) = \beta R^F E_t[u'(c_{t+1} - X_{t+1})]$. This implies that $E_t[u'(c_{t+1} - X_{t+1})] = (\beta R^F)^{-1} u'(c_t - X_t)$ and $MU_t^{INT} = \beta' u'(c_t - X_t) - b\beta' \sum_{s=1}^S \beta^s \phi^s u'(c_{t+s} - X_{t+1}) = const. \times \beta' u'(c_t - X_t) = const. \times MU_t^{EXT}$. Therefore, if the Euler equations with external habit are satisfied, so are the Euler equations with internal habit.

sum of past consumption levels, as in equation (2.2). The non-linear specification of the external habit is reverse-engineered such that the risk free rate is either constant or has low variability. The model has in common with the Constantinides (1990) linear internal habit model the intuition that the RRA coefficient is an increasing function of the habit-to-consumption ratio and is, therefore, countercyclical.

The exogenous specification of the non-linear dependence of the external habit on past changes in consumption affords Campbell and Cochrane (1999) the flexibility to generate predictable long-horizon returns and realistic behavior of several macroeconomic series along the business cycle. Their model hits the aggregate price-dividend ratio in a number of periods but misses it in the 1950s and 1990s. The major contribution of this model is that it sets a broad and ambitious research agenda that goes beyond the equity premium puzzle and addresses the broad interface between finance and macroeconomics. A drawback of the model is that the implied RRA coefficient is implausibly high and highly volatile: it takes the value of eighty in the steady state and a much higher value in economic recessions. Nevertheless, the model provides a flexible theoretical tool for exploring a broad set of issues, leaving concerns about the high risk aversion as a topic for future research.

Menzly, Santos, and Veronesi (2004) developed a theoretical model that combines the Campbell and Cochrane preferences with an interesting cash flow process across industries. Dividend *shares* across industries vary over time in a stationary way, so no one industry dominates over time. The model explains two puzzling stylized empirical observations: first, dividend yield forecasts stock returns but the strength of this relationship varies across subperiods and is absent in some subperiods; second, even though dividend growth is forecastable, dividend yield does not forecast dividend growth in simple regressions. The key to their model is that an increase in expected dividend growth decreases the dividend yield *and* increases the expected return. This weakens the commonly explored predictability of expected return by the dividend yield. Menzly, Santos, and Veronesi (2004) reported that their version of the non-linear external habit model helps explain the cross-section of asset returns.

Li (2001) reported that in both the linear and non-linear external habit model, the habit-to-consumption ratio has limited success in explaining the time series of the premia

of stock and bond portfolios. Korniotis (2005) reported positive empirical results in testing factor pricing models that recognize non-linear external habit but only in conjunction with the assumption that the market is incomplete.

High on the agenda for future research is the development of habit models that combine the low implied risk aversion of the Constantinides (1990) linear internal habit model with the flexibility of the nonlinear external habit model of Campbell and Cochrane (1999). On the empirical front, it is important to investigate the extent to which the empirical success of forward-looking consumption growth in explaining the cross-section of asset returns is due to internal habit.

3. Limited stock market participation and per capita consumption

The low covariance of the growth rate of per capita consumption with equity returns is a major stumbling block in explaining the mean aggregate equity premium and the cross-section of asset returns, in the context of a representative-consumer economy with time separable preferences. The incorrect specification of per capita consumption contributes to the low correlation of the equity return with the growth rate of per capita consumption. Failure to adjust for the limited participation of households in the stock market is one source of this misspecification.

Whereas we understood all along that many households whose consumption is counted in the NIPA-reported aggregate capital consumption do not hold equities, it took a paper by Mankiw and Zeldes (1991) to point out that the emperor has no clothes.²¹ Even though 52 percent of the United States adult population held stock either directly or indirectly in 1998, compared to 36 percent in 1989, stockholdings remain extremely concentrated in the hands of the wealthiest few.²² Furthermore, wealthy entrepreneurs may be infra marginal in the stock market if their wealth is tied up in private equity.

In Section 1, I pointed out that the unconditional mean equity premium is equal to the product of the RRA coefficient of the representative consumer and the covariance of per capita consumption growth with the market return. I now explicitly distinguish between the NIPA-reported aggregate consumption, C_t^{NIPA} , and the aggregate consumption, $I \times C_t \equiv \sum_{i=1}^I c_t^i$, by the subset of households, $i, i = 1, \dots, I$, who are

²¹ Since then, several papers have studied the savings and portfolio composition of households, stratified by income, wealth, age, education, and nationality. See Blume and Zeldes (1993), Haliassos and Bertaut (1995), Heaton and Lucas (2000), Poterba (2001), and the collected essays in Guiso, Haliassos, and Jappelli (2001).

²² The percentages are only indicative of the extent of non-participation for two reasons. First, small stockholders may still be marginal. Second, given the wide variety of retirement plans, households have widely varying control on their investment in these plans.

marginal in the equities market, identified as the stockholders. The NIPA-reported aggregate consumption growth, $C_t^{NIPA} / C_{t+1}^{NIPA}$, is a noisy measure of the stockholders' aggregate consumption growth, C_t / C_{t+1} . By explicitly recognizing the limited stock market participation of households, I may coax the theory to predict a higher value of the mean equity premium without increasing the RRA coefficient and without predicting an implausibly high value for the risk free rate.

Mankiw and Zeldes (1991) calculated per capita food consumption of a subset of households, designated as asset holders according to a criterion of asset holdings above some threshold. They found that the implied RRA coefficient decreases as the threshold is raised. Brav and Geczy (1995) confirmed their result by using per capita consumption of non-durables and services, reconstructed from the Consumer Expenditure Survey (CEX) database. Attanasio, Banks, and Tanner (2002), Brav, Constantinides, and Geczy (2002), and Vissing-Jorgensen (2002) found some evidence that per capita consumption growth can explain the equity premium with a relatively high value of the RRA coefficient, once we account for limited stock market participation. However, Brav, Constantinides, and Geczy (2002) pointed out that the statistical evidence is weak and the results are sensitive to the experimental design. My interpretation of these empirical results is that recognition of limited stock market participation is a step in the right direction but this step alone is insufficient to explain asset returns.

Constantinides, Donaldson, and Mehra (2002) introduced borrowing constraints in an overlapping-generations exchange economy and provided an endogenous explanation for the limited participation of young consumers in the stock market. They also demonstrated that limited participation in the borrowing constrained economy works in the right direction in the increasing the equity premium, lowering the bond yield and lowering the variance of the bond and equity returns.

In the Constantinides et al (2002) economy, consumers live for three periods. In the first period, a period of human capital acquisition, consumers receive a relatively low endowment income. In the second period, consumers are employed and receive wage income subject to large uncertainty. In the third period, consumers retire and consume the assets accumulated in the second period. The key feature is that the bulk of the *future* income of the young consumers is derived from their wages forthcoming in their middle

age, while the *future* income of the middle-aged consumers is derived primarily from their savings in equity and bonds.

The young would like to invest in equity, given the observed large equity premium. However, they are unwilling to decrease their current consumption in order to save by investing in equity, because the bulk of their lifetime income is derived from their wages forthcoming in their middle age. They would like to borrow, but the borrowing constraint prevents them from doing so. Human capital alone does not collateralize major loans in modern economies for reasons of moral hazard and adverse selection. The model explains why many consumers do not participate in the stock market in the early phase of their life cycle.

The future income of the middle-aged consumers is derived from their current savings in equity and bonds. Therefore, the risk of holding equity and bonds is concentrated in the hands of the middle-aged saving consumers. This concentration of risk generates the high equity premium and the demand for bonds, in addition to the demand for equity, by the middle-aged.²³ The model recognizes and addresses simultaneously, at least in part, the equity premium, the limited participation in the stock market, and the demand for bonds.

Constantinides et al (2002) calibrated the model to match the following eight targets: the average share of income going to labor; the average share of income going to the labor of the young; the average share of income going to interest on government debt; the coefficient of variation of the 20-year wage income of the middle aged; the coefficient of variation of the 20-year aggregate income; the 20-year autocorrelation of the labor income; the 20-year autocorrelation of the aggregate income; and the 20-year cross-correlation of the labor income and the aggregate income. Since the length of one period in this model is twenty years, for all securities (equity, bond or consol), the annualized mean and standard deviation of the return are defined as the mean and

²³ See also the discussion in the related papers by Bertaut and Haliassos (1997), Bodie, Merton, and Samuelson (1992), Jagannathan and Kocherlakota (1996) and Storesletten, Telmer, and Yaron (2001).

standard deviation, respectively, of the annualized 20-year holding period return. Below, I reproduce the first panel of Table 1 in Constantinides et.al. (2002).²⁴

	Borrowing-constrained	Borrowing-unconstrained
Mean equity return	8.4	10.2
Std dev of equity return	23.0	42.0
Mean bond return	5.1	9.0
Std dev of bond return	15.4	27.6
Mean premium over bond	3.4	1.1
Std dev of premium over bond	18.4	31.6
Mean consol return	3.7	9.9
Std dev of consol return	19.1	27.6
Mean premium over consol	4.7	0.3
Std dev of premium over consol	10.5	5.2
Corr of wages and dividend	-0.43	-0.43
Corr of wages and premium	-0.02	0.00

The borrowing constraint decreases the mean return of the 20-year or consol bond by about a factor of two. This observation is robust to the calibration of the correlation and auto-correlation of the labor income of the middle-aged with the aggregate income. The borrowing constraint goes a long way, albeit not all the way, towards resolving the risk free rate puzzle. If the young were able to borrow, they would do so and purchase equity; the borrowing activity of the young would raise the bond return, thereby exacerbating the risk-free rate puzzle.

Second, the mean equity premium over the 20-year or consol bond is about 4%. This is satisfactory given that long term bond returns typically command a premium over the short term risk free rate. This premium drastically decreases when the borrowing constraint is relaxed. If the young were able to borrow, the increase in the bond return would induce the middle-aged to shift their portfolio holdings in favor of bonds; the increase in the demand for equity by the young and the decrease in the demand for equity by the middle-aged work in opposite directions; on balance, the effect would be to increase the return on both equity and the bond while simultaneously shrinking the equity premium.

²⁴ The RRA coefficient is set at 6; the coefficient of variation of the (20-year) aggregate income is set at 0.20; the coefficient of variation of the (20-year) wages is set at 0.25; the auto-correlation of aggregate income, the auto-correlation of wages, and the correlation of aggregate income and wages are all set at 0.1. The consol bond is in positive net supply and the one-period (20-year) bond is in zero net supply.

Third, the correlation of the labor income of the middle-aged and the equity premium over the 20-year bond is much smaller in absolute value than the exogenously-imposed correlation of the labor income of the middle-aged and the dividend. Thus, equity is attractive to the young because of the large mean equity premium *and* the low correlation of the premium with the wage income of the middle-aged, thereby corroborating another important dimension of the model. In equilibrium, it turns out that the correlation of the wage income of the middle-aged and the equity return is low.²⁵ The young consumers would like to invest in equity because equity return has low correlation with their future consumption, if their future consumption is derived from their future wage income. However, the borrowing constraint prevents them from purchasing equity on margin. Furthermore, since the young consumers are relatively poor and have an incentive to smooth their intertemporal consumption, they are unwilling to decrease their current consumption in order to save by investing in equity. Therefore, the young choose not participate in the equity market.

Finally, the borrowing constraint lowers the standard deviation of the annualized, 20-year equity and bond returns. However, the standard deviation of the equity return and particularly the standard deviation of the bond return remain high.

In related work, Guvenen (2005) calibrated a general equilibrium model with two classes of infinitely-lived agents, one that are exogenously forbidden from investing in equities and other that hold the entire stock of equity. The model replicates many features of the macroeconomy and the real business cycle through the channel of unrealistically high variance of the consumption growth of the stockholders.

Constantinides, Donaldson, and Mehra (2002, 2005) addressed yet another major misspecification of per capita consumption. They included *inter vivos* gifts and *post mortem* bequests in the definition of consumption thereby increasing the covariance of equity returns with the growth rate of consumption. In the context of an overlapping-generations economy, they modeled the “consumption” of the old households as consisting of the “joy of giving”, c_B , in the form of *inter vivos* gifts and *post mortem*

²⁵ The low correlation of the wage income of the middle-aged and the equity return is a property of the equilibrium and obtains for a wide range of values of the assumed correlation of the wage income of the middle-aged and the dividend.

bequests, in addition to the direct consumption, c_D , of durable goods, services, and the service flow from non-durables goods.

The direct consumption of the old cohort, wealthy enough to be non-trivial investors in the capital markets is dictated by the state of their health and their medical expenses rather than fluctuations of their wealth. The correlation between the direct consumption of the old and the stock market return is *low*, a prediction that is borne out empirically. Constantinides et al (2005) captured this feature by assuming that direct consumption is held constant because of a binding constraint on direct consumption. The bequest equals the wealth of the old, net of the direct consumption of the old, and is *a fortiori* highly correlated with the stock market return. Thus, in an economy that recognizes bequests, the marginal rate of substitution with respect to the bequests of the old, is better able to explain the mean equity premium and the risk free rate than the marginal rate of substitution with respect to the direct consumption of the old.

How does the OLG economy with bequests as in Constantinides et al (2005) differ in its pricing implications from the OLG economy without bequests as in Constantinides et al (2002), if the wage and dividend processes in the two economies are identical? With low risk aversion, the calibrated model produces realistic values for the mean equity premium and the risk free rate, the variance of the equity premium, and the ratio of bequests to wealth. However, the variance of the risk free rate is unrealistically high. Security prices tend to be substantially higher in an economy with bequests as compared to an otherwise identical one where bequests are absent. We are able to keep the prices low and generate reasonable returns and the equity premium by stipulating that a portion of the bequests skips a generation and is received by the young, rather than by the middle aged.

Whereas borrowing constraints may explain the non-participation of young consumers in the financial markets, we are still far from a satisfactory explanation of the limited participation of middle-aged consumers in the financial markets. A better understanding of limited participation may aid our understanding of the equity premium puzzle and the returns of financial assets. Ang, Bekaert and Liu (2005) took a step in this direction by demonstrating that investors endowed with “disappointment aversion”

preferences as in Gul (1991) may choose to refrain from participating in some financial markets for a wide range of preference parameters.

4. Incomplete markets and idiosyncratic income shocks

In economic recessions, investors are exposed to the double hazard of stock market losses and job loss. Investment in equities not only fails to hedge the risk of job loss but accentuates its implications. Investors require a hefty equity premium in order to be induced to hold equities.

The argument hinges on the fact that the market for job-loss insurance is incomplete. Moral hazard impedes the development of unemployment insurance markets and extant unemployment compensation provides inadequate protection for investors wealthy enough to be significant players in the financial markets. For example, with job-loss insurance, an increase of one percent in the probability of job loss merely manifests itself as a one percent decrease in per capita consumption growth—a risk which is too small to generate the observed equity premium. Job loss is but one example of idiosyncratic income shocks that have the potential to generate a sizable equity premium.

The observed correlation of per capita consumption growth with stock returns is low. Over the years, I have grown skeptical of how meaningful an economic construct per capita consumption is, and how hard we should push per capita consumption to explain returns. Per capita consumption is a meaningful economic concept only if the market is complete or all consumers have quadratic utility. Any time that we model the household sector by a representative consumer and proxy consumption with per capita (or, aggregate) consumption, we suppress the potentially major economic impact of uninsurable idiosyncratic income shocks.

I begin to formalize these ideas with the observation that, in the competitive equilibrium of a complete market economy, heterogeneous households allocate their consumption in a way that they equalize, state by state, their marginal rate of substitution. Negishi (1960) pointed out that the same consumption allocation is attained when a social planner maximizes the judiciously weighted sum of the households' utility functions. For this result to hold, the weight of each household's utility function is chosen to be

inversely proportional to the Lagrange multiplier of the budget constraint of the respective household. The claim follows from the observation that the households' first-order conditions are identical to the first-order conditions of a social planner.

Constantinides (1982) further showed that the social planner's problem can be split into two sub-problems.²⁶ In the first sub-problem, one constructs an increasing and concave utility function out of the social planner's weighted sum of the households' utility functions. In the second sub-problem, the social planner maximizes the above-constructed utility function and ends up allocating consumption at each state equal to the total consumption allocated at that state in the heterogeneous-household economy. This implies that the equilibrium in a heterogeneous-household, full-information economy is isomorphic in its pricing implications to the equilibrium in a representative-household, full-information economy, if households have von Neumann-Morgenstern preferences. The strong assumption of market completeness is indirectly built into asset pricing models in finance and neoclassical macroeconomic models through the assumption of the existence of a representative household.

Bewley (1982), Mankiw (1986), and Mehra and Prescott (1985) suggested the potential of enriching the asset-pricing implications of the representative-household paradigm, by relaxing the assumption of complete markets.²⁷ Constantinides and Duffie (1996) found that incomplete markets substantially enrich the implications of the representative-household model. Their main result is a proposition demonstrating, by construction, the existence of household income processes, consistent with given aggregate income and dividend processes, such that equilibrium equity and bond price processes match the given equity and bond price processes.

The starting point in the Constantinides and Duffie (1996) theory, are the $I \times J$ Euler equations (1.4) of individual consumption by the I consumers for the J assets, specialized for preferences that imply constant relative risk aversion:

$$E_t \left[\beta g_{i,t+1}^{-A} R_{j,t+1} \right] = 1, \quad i = 1, \dots, I; \quad j = 1, \dots, J. \quad (4.1)$$

²⁶ See also the discussion in Mehra and Prescott (1980).

²⁷ There is an extensive literature on the hypothesis of complete consumption insurance. See Altonji, Hayashi and Kotlikoff (1992), Attanasio and Davis (1997), Cochrane (1991), and Mace (1991).

The consumption growth of the i^{th} consumer is $g_{i,t+1} \equiv c_{t+1}^i / c_t^i$ and the RRA coefficient is A . The point of departure from standard theory is the recognition that market incompleteness rules out the step of replacing the Euler equations (4.1) by the Euler equation (1.6) of the representative consumer.

A stochastic discount factor (SDF), or pricing kernel, is defined as any random variable m_{t+1} with the following property:

$$E_t [m_{t+1} R_{j,t+1}] = 1, \quad j = 1, \dots, J. \quad (4.2)$$

Therefore, equation (4.1) states that each consumer's marginal rate of substitution, $\beta g_{i,t+1}^{-A}$, is a valid SDF.

I sum the $I \times J$ Euler equations (4.1) across households and obtain the expression

$$E_t \left[\beta \left\{ I^{-1} \sum_{i=1}^I g_{i,t+1}^{-A} \right\} R_{j,t+1} \right] = 1, \quad j = 1, \dots, J. \quad (4.3)$$

Equation (4.3) states that the weighted sum of the consumers' marginal rate of substitution is a valid SDF also:

$$m_{t+1} = \beta \left\{ I^{-1} \sum_{i=1}^I g_{i,t+1}^{-A} \right\}. \quad (4.4)$$

I expand equation (4.4) as a Taylor series up to cubic terms and obtain the following approximation for the SDF:

$$m_{t+1} = \beta g_{t+1}^{-A} \left\{ 1 + \frac{A(A+1)}{2} I^{-1} \sum_{i=1}^I \left(\frac{g_{i,t+1}}{g_{t+1}} - 1 \right)^2 - \frac{A(A+1)(A+2)}{6} I^{-1} \sum_{i=1}^I \left(\frac{g_{i,t+1}}{g_{t+1}} - 1 \right)^3 \right\}. \quad (4.5)$$

The term $g_{t+1} \equiv I^{-1} \sum_{i=1}^I g_{i,t+1}$ is the sample mean of the consumption growth rate across consumers; $I^{-1} \sum_{i=1}^I \left(\frac{g_{i,t+1}}{g_{t+1}} - 1 \right)^2$ is the squared coefficient of variation of the consumption growth rate across consumers; and $I^{-1} \sum_{i=1}^I \left(\frac{g_{i,t+1}}{g_{t+1}} - 1 \right)^3$ is a measure of the skewness of the cross sectional variation of consumption growth.

The theory requires that the idiosyncratic income shocks have three properties in order to explain the returns on financial assets. First, they must be *uninsurable*. If there exist a complete set of markets, the equilibrium of a heterogeneous-household, full-information economy is isomorphic in its pricing implications to the equilibrium of a representative-household, full-information economy and household consumption growth cannot do better than aggregate consumption growth in explaining the returns. To see this, note that, if a complete set of markets exists, then the heterogeneous households are able to equalize, state by state, their marginal rates of substitution. In particular for any state s at time $t + 1$, there exists a state-specific but consumer-independent parameter λ_s such that $g_{i,t} = g_t = \lambda_s$, $i = 1, \dots, I$. The SDF in equations (4.5) simplifies into

$$m_{t+1} = \beta g_{t+1}^{-A} \quad (4.6)$$

or, equivalently, into²⁸

$$m_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-A}. \quad (4.7)$$

²⁸ When there is a complete set of markets, equation (4.7) follows from the fact that, for any state s ,

$$\frac{C_{t+1}}{C_t} = \frac{I^{-1} \sum_{i=1}^I c_{t+1}^i}{I^{-1} \sum_{i=1}^I c_t^i} = \lambda_s = g_{t+1}.$$

I expect that the SDF given by equation (4.7) is less susceptible to observation error than the SDF given by equation (4.6).

Therefore, if there exist a complete set of markets, the equilibrium of a heterogeneous-household, full-information economy is isomorphic in its pricing implications to the equilibrium of a representative-household, full-information economy and household consumption growth cannot do better than aggregate consumption growth in explaining the returns.

Second, the theory requires that the idiosyncratic income shocks be *persistent*. If the shocks are transient, then households can smooth their consumption by borrowing or by drawing down their savings.²⁹

Third, the conditional variance, or some higher moment of the income shocks, must be counter-cyclical. In equation (4.5), the SDF is monotone increasing in the conditional variance. If the conditional variance is counter-cyclical, then the SDF is counter-cyclical and co-varies negatively with the market return, even though aggregate consumption has low covariance with the market return. In principle, this negative co-variation gives rise to an equity premium that is absent in a complete market. Even if the conditional variance is not counter-cyclical, the cyclical or counter-cyclical behavior of some higher moment of the income shocks may generate a high equity premium. For example, in equation (4.5), the SDF is monotone decreasing in the conditional skewness. If the conditional skewness is cyclical, then the SDF is counter-cyclical and co-varies negatively with the market return, giving rise to an equity premium.

A good example of a major uninsurable income shock is job loss. Job loss is *uninsurable* because unemployment compensation is inadequate. Layoffs have *persistent* implications on household income, even though the laid-off workers typically find another job quickly.³⁰ Layoffs are *counter-cyclical* as they are more likely to occur in recessions.

²⁹ Aiyagari and Gertler (1991) and Heaton and Lucas (1996) found that consumers facing *transient* shocks come close to the complete-markets rule of complete risk sharing even with transaction costs and/or borrowing costs, *provided that the supply of bonds is not restricted to an unrealistically low level*.

³⁰ The empirical evidence is sensitive to the model specification. Heaton and Lucas (1996) modeled the income process as *univariate* and provided empirical evidence from the Panel Study on Income Dynamics (PSID) that the idiosyncratic income shocks are transitory. Storesletten, Telmer, and Yaron (2001) modeled the income process as *bivariate* and provided empirical evidence from the PSID that the

Recall that the main result in Constantinides and Duffie (1996) is a proposition demonstrating, by construction, the existence of household income processes, consistent with given aggregate income and dividend processes, such that equilibrium equity and bond price processes match the given equity and bond price processes. The proposition implies that the Euler equations of household (but not necessarily of per capita) consumption must hold. Furthermore, since the given price processes have embedded in them whatever predictability of returns by the price-dividend ratios, dividend growth rates, and other instruments that the researcher cares to ascribe to returns, the equilibrium price processes have this predictability built into them by construction.

The first implication of the theory is an explanation of the counter-cyclical behavior of the equity risk premium: the risk premium is highest in a recession because the stock is a poor hedge against the uninsurable income shocks, such as job loss, that are more likely to arrive during a recession.

The second implication is an explanation of the unconditional equity premium puzzle: even though per capita consumption growth is poorly correlated with stocks returns, investors require a hefty premium to hold stocks over short-term bonds because stocks perform poorly in recessions, when the investor is most likely to be laid off.

In principle, I may directly test the $I \times J$ system of Euler equations (4.1) of household consumption. In the US, the best available disaggregated consumption data are provided by the department of Labor Statistics' Consumer Expenditure Survey (CEX) of quarterly consumption of selected households (not individual consumers). In practice the direct test is difficult because household consumption data are reported with substantial error.

Brav, Constantinides, and Geczy (2002) provided empirical evidence on the importance of uninsurable idiosyncratic income risk on pricing. They put forth a series of candidate SDFs: the marginal rate of substitution of the representative consumer as in equations (4.6) or (4.7); the Taylor series expansion of the SDF up to quadratic terms; the Taylor series expansion of the SDF up to cubic terms, as in equation (4.5); and a log-

idiosyncratic income shocks have a highly persistent component that becomes more volatile during economic contractions. Storesletten, Telmer, and Yaron (2000) corroborated the latter evidence by studying household consumption over the life cycle.

linearized expansion of the SDF. They estimated the RRA coefficient and tested the set of Euler equations of household consumption on the premium of the value-weighted and the equally weighted market portfolio return over the risk free rate and on the premium of value stocks over growth stocks.³¹

Brav et al (2002) did not reject the Euler equations of household consumption with RRA coefficient between two and four when the candidate SDF is the Taylor series expansion of the SDF up to cubic terms, as in equation (4.5). A RRA coefficient between two and four is economically plausible. They rejected the Euler equations of household consumption with any value of the RRA coefficient when the Taylor series expansion of the SDF does not include the cubic terms. This implies that, in addition to the mean and variance, the *skewness* of the cross-sectional distribution is important in explaining the equity premium. They also found that the log-linearized expansion of the SDF fails to explain the premia, possibly because the log-linearization downplays the effect of outliers. These results emphasize the role of the higher moments of the cross-sectional distribution in explaining the premia.

Krebs (2002) provided a theoretical justification as to why it is possible that neither the variance nor the skewness, but higher moments of the cross-sectional distribution are important in explaining the equity premium. He extended the Constantinides and Duffie (1996) model by introducing rare idiosyncratic income shocks that drive consumption close to zero. In his model, the conditional variance and skewness of the idiosyncratic income shocks are nearly constant over time. Despite this, Krebs demonstrated that the original proposition of Constantinides and Duffie remains valid, that is, there exist household income processes, consistent with given aggregate income and dividend processes, such that equilibrium equity and bond price processes match the given equity and bond price processes. Essentially, he provided a theoretical justification as to why it may be hard to empirically detect the rare but catastrophic shocks in the low-order cross-sectional moments of household consumption growth.

³¹ In related studies, Jacobs (1999) studied the PSID database on food consumption; Cogley (2002) and Vissing-Jorgensen (2002) studied the CEX database on broad measures of consumption; Jacobs and Wang (2004) studied the CEX database by constructing synthetic cohorts; and Ait-Sahalia, Parker, and Yogo (2004) proxied household consumption with the purchases of certain luxury goods.

Jacobs and Wang (2004) expanded the set test assets employed by Brav et al (2002) to the set of the 25 Fama-French size and value portfolios and provided additional empirical evidence on the importance of uninsurable idiosyncratic income risk on pricing. They found that a two-factor asset pricing model, with the mean and cross-sectional variance of the household consumption growth rate as factors, significantly outperforms the capital asset pricing model in explaining the cross-section of asset returns. See also Semenov (2005) for a discussion of the role of higher moments.

Korniotis (2005) investigated the potential incompleteness of the market across U.S. states while assuming complete consumption insurance within each state. He tested the Euler equation of consumption for each state, where state consumption is proxied by the annual state-wide sales at retail establishments. He found that a factor pricing model that recognizes this incompleteness with the cross-sectional variance of the state consumption as a factor has modest success in explaining the cross-section of asset returns. He also found that a factor pricing model that simultaneously recognizes this incompleteness and non-linear external habit better explains the cross-section of asset returns.

5. Concluding remarks

I examined the observed asset returns and conclude that the evidence does not support the case for abandoning the rational economic model. I argued that the standard model is greatly enhanced by relaxing some of its assumptions. In particular, I argued that we go a long way toward addressing market behavior by recognizing that consumers face uninsurable and idiosyncratic income shocks, for example, the loss of employment. The prospect of such events is higher in economic downturns and this observation takes us a long way toward understanding both the unconditional moments of asset returns and their variation along the business cycle.

I also argued that we should account for the limited stock market participation by households in constructing per capita consumption growth rate by excluding from the measure of per capita consumption the households that are infra-marginal in the stock market. In this context, life cycle considerations are important and often overlooked in Finance. Borrowing constraints become important when placed in the context of the life cycle. The fictitious representative consumer that holds all the stock market and bond market wealth does not face credible borrowing constraints. Young consumers, however, do face credible borrowing constraints. I traced their impact on the equity premium, the demand for bonds—who holds bonds if the equity premium is so high? —and on the limited participation of consumers in the capital markets.

Finally, I argued that relaxing the assumption of convenience that preferences are time separable drives a wedge between the preference properties of risk aversion and intertemporal elasticity of substitution, within the class of von Neumann Morgenstern preferences. Further work along these lines may enhance our understanding of price behavior along the business cycle with credibly low risk aversion coefficient.

Has the equity premium puzzle been resolved? I don't think so. However, the pursuit of an explanation has helped us broaden the scope of our investigation in several important ways. We are now interested in understanding the mean, higher moments, co-variability, and predictability of the return of different classes of financial assets. At the macro level, we study the short-term risk-free rate, the term premium of long-term bonds over the risk-free rate, and the aggregate equity premium of the stock market over the

risk-free rate. At the micro level, we study the premium of individual stock returns and of classes of stocks, such as the small-capitalization versus large-capitalization stocks, the “value” versus “growth” stocks, and the historical record of losing versus winning stocks.

Our pursuit has led us to study a broader class of preferences, beliefs and learning mechanisms, the evolution of state variables, market incompleteness, market imperfections, and notions of liquidity than those embedded in the standard neoclassical model. Our pursuit has also helped us gain a better understanding of data problems such as limited participation of consumers in the stock market; temporal aggregation and survival biases. In my eclectic discussion of these issues, I hope to have convinced the reader that the integration of the notions of *habit persistence*, *incomplete markets*, the *life cycle*, *borrowing constraints*, and other sources of *limited stock market participation* is a promising vantage point from which to study the prices of asset and their returns both theoretically and empirically.

REFERENCES

- Abel, A. B. (1990), Asset prices under habit formation and catching up with the Joneses, *American Economic Review Papers and Proceedings* 80: 38-42.
- Ait-Sahalia, Y., J. A. Parker and M. Yogo (2004), Luxury goods and the equity premium, *Journal of Finance* 59: 2959-3004.
- Aiyagari, R. S. and M. Gertler (1991), Asset returns with transactions costs and uninsured individual risk, *Journal of Monetary Economics* 27: 311-331.
- Altonji, J. G., F. Hayashi and L. J. Kotlikoff (1992), Is the extended family altruistically linked? *American Economic Review* 82: 1177-1198.
- Alvarez, F. and U. J. Jermann (2001), Quantitative asset pricing implications of endogenous solvency constraints, *Review of Financial Studies* 14: 1117-1152.
- Ang, A., G. Bekaert and J. Liu (2005), Why stocks may disappoint, *Journal of Financial Economics* 76: 471-508.
- Attanasio, O. P., J. Banks and S. Tanner (2002), Asset holding and consumption volatility, *Journal of Political Economy* 110: 771-792.
- Attanasio, O. P. and S. J. Davis (1997), Relative wage movements and the distribution of consumption, *Journal of Political Economy* 104: 1227-1262.
- Bakshi, G. and Z. Chen (1996), The spirit of capitalism and stock market prices, *American Economic Review* 86: 133-157.
- Bansal, R. and J. W. Coleman (1996), A monetary explanation of the equity premium, term premium, and risk-free rate puzzles, *Journal of Political Economy* 104: 1135-1171.
- Bansal, R., R. F. Dittmar and C. T. Lundblad (2004), Consumption, dividends and the cross-section of equity returns, *Journal of Finance* (forthcoming).
- Bansal, R. and A. Yaron (2004), Risks for the long run: a potential resolution of asset pricing puzzles, *Journal of Finance* 59: 1481-1509.
- Benartzi, S. and R. H. Thaler (1995) Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110: 73-92.
- Barro, R. J. (2005), Rare events and the equity premium, Working paper, Harvard University.

- Bertaut, C. C. and M. Haliassos (1997), Precautionary portfolio behavior from a life-cycle perspective, *Journal of Economic Dynamics and Control* 21: 1511-1542.
- Bewley, T. F. (1982), Thoughts on tests of the intertemporal asset pricing model, Working paper, Northwestern University.
- Blume, M. E. and S. P. Zeldes (1993), The structure of stock ownership in the U.S., Working paper, University of Pennsylvania.
- Bodie, Z., R. C. Merton and W. F. Samuelson (1992), Labor supply flexibility and portfolio choice in a life cycle model, *Journal of Economic Dynamics and Control* 16: 427-449.
- Boldrin, M., L. J. Christiano and J. D. M. Fisher (2001), Habit persistence, asset returns, and the business cycle, *American Economic Review* 91: 149-166.
- Braun, P. A., G. M. Constantinides and W. E. Ferson (1993), Time Nonseparability in Aggregate Consumption: International Evidence, *European Economic Review* 37: 897-920.
- Brav, A., and C. C. Geczy (2002), Asset pricing with heterogeneous consumers and limited participation: Empirical evidence, *Journal of Political Economy* 110: 793-824.
- Brav, A. and C. C. Geczy (1995), An empirical resurrection of the simple consumption CAPM with power utility, Working paper, University of Chicago.
- Breeden, D. T. (1979), An intertemporal asset pricing model with stochastic consumption and investment opportunities, *Journal of Financial Economics* 7: 265-296.
- Brown, S. J., W. N. Goetzmann and S. Ross (1995), Survival, *Journal of Finance* 50: 853-873.
- Cagetti, M., L. P. Hansen, T. Sargent and N. Williams (2002), Robustness and pricing with uncertain growth, *Review of Financial Studies* 15: 363-404.
- Campbell, J. Y., (2001), Asset pricing at the millennium, *Journal of Finance* 55: 1515-1567.
- Campbell, J. Y. (2003), Consumption-based asset pricing, in: G. M. Constantinides, M. Harris and R. Stulz, eds., *Financial Markets and Asset Pricing: Handbook of the Economics of Finance*, Volume 1B, Handbooks in Economics 21 (North-Holland, Amsterdam) 803-887.

- Campbell, J. Y. and J. H. Cochrane (1999), By force of habit: a consumption-based explanation of aggregate stock market behavior, *Journal of Political Economy* 107: 205-251.
- Campbell, J. Y., A. W. Lo and A. C. MacKinlay (1997), *The Econometrics of Financial Markets* (Princeton University Press, Princeton, NJ).
- Cecchetti, S. G., P. Lam and N. Mark (1994), The equity premium and the risk free rate: matching the moments, *Journal of Monetary Economics*, 31: 21-46.
- Cecchetti, S. G., P. Lam and N. Mark (2000), Asset pricing with distorted beliefs: are equity returns too good to be true?, *American Economic Review* 90: 787-805.
- Chen, X. and S. C. Ludvigson (2005), Land of addicts? An empirical investigation of habit-based asset pricing models, Working paper, New York University.
- Cochrane, J. H. (1991), A simple test of consumption insurance, *Journal of Political Economy* 99: 957-976.
- Cochrane, J. H. (1997), Where is the market going? Uncertain facts and novel theories, *Economic Perspectives* (Federal Reserve Bank of Chicago) 21: 3-37.
- Cochrane, J. H. (2001), *Asset Pricing* (Princeton University Press, Princeton, NJ).
- Cochrane, J. H. and L. P. Hansen (1992), Asset pricing explorations for macroeconomics, in O. J. Blanchard and S. Fischer, eds., NBER, *Macroeconomics Annual* (MIT Press, Cambridge, MA).
- Cogley, T. (2002), idiosyncratic risk and the equity premium: evidence from the consumer expenditure survey, *Journal of Monetary Economics* 49: 309-334.
- Constantinides, George M., 1982, Intertemporal asset pricing with heterogeneous consumers and without demand aggregation, *Journal of Business* 55: 253-267.
- Constantinides, G. M. (1990), Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98: 519-543.
- Constantinides, G. M. (2002), Rational asset prices, *Journal of Finance* 57, 1567-1591.
- Constantinides, G. M., J. B. Donaldson and R. Mehra (2002), Junior can't borrow: a new perspective on the equity premium puzzle, *Quarterly Journal of Economics* 117: 269-296.
- Constantinides, G. M., J. B. Donaldson and R. Mehra (2005), Junior is rich: bequests as consumption, Working paper, University of Chicago.

- Constantinides, G. M. and D. Duffie (1996), Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104: 219-240.
- Daniel, K. and D. Marshall (1997), The equity premium puzzle and the risk-free rate puzzle at long horizons, *Macroeconomic Dynamics* 1: 452-484.
- Danthine, J.-P. and J. B. Donaldson (1999), Non-falsified expectations and general equilibrium asset pricing: the power of the peso, *Economic Journal* 109: 607-635.
- Danthine, J.-P., J. B. Donaldson and R. Mehra (1992), The equity premium and the allocation of income risk, *Journal of Economic Dynamics and Control* 16: 509-32.
- De Santis, M. (2004), Interpreting aggregate stock market behavior: how far can the standard model go?, Working paper, UC Davis.
- Duesenberry, J. S. (1949), *Income, Saving, and the Theory of Consumer Behavior* (Harvard University Press, Cambridge, MA).
- Epstein, L. G. and S. E. Zin (1991), Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis, *Journal of Political Economy* 99: 263-286.
- Ferson, W. E. and G. M. Constantinides (1991), Habit persistence and durability in aggregate consumption, *Journal of Financial Economics* 29: 199-240.
- Ferson, W. E. and C. R. Harvey (1992), Seasonality and consumption-based asset pricing, *Journal of Finance* 47: 511-552.
- Friend, I. and M. E. Blume (1975), The demand for risky assets, *The American Economic Review* 65: 900-922.
- Gabaix, X. and D. Laibson (2001), The 6D bias and the equity premium puzzle, in Ben Bernanke and Ken Rogoff, eds., *NBER Macroeconomics Annual 2001* (MIT Press, Cambridge, MA).
- Gallant, A. R., L. P. Hansen and G. Tauchen (1990), Using conditional moments of asset payoffs to infer the volatility of intertemporal marginal rates of substitution, *Journal of Econometrics* 45: 141-179.
- Grossman, S. J. and R. J. Shiller (1981), The determinants of the variability of stock market prices, *American Economic Review Papers and Proceedings* 71: 222-227.

- Guidolin, M. (2004), Pessimistic beliefs under rational learning: quantitative implications for the equity premium puzzle, Working paper, University of Virginia.
- Guiso, L., M. Haliassos and T. Jappelli (2001), *Household Portfolios* (MIT Press, Cambridge, MA).
- Gul, F. (1991), A theory of disappointment aversion, *Econometrica* 59: 667-686.
- Guvenen, F. (2005), A parsimonious macroeconomic model for asset pricing: habit formation or cross-sectional heterogeneity, Working paper, University of Rochester.
- Haliassos, M. and C. C. Bertaut (1995), Why do so few hold stocks? *The Economic Journal* 105: 1110-1129.
- Hansen, L. P., J. C. Heaton and N. Li (2005), Consumption strikes back? Measuring long run risk, Working paper, University of Chicago.
- Hansen, L. P. and R. Jagannathan (1991), Implications of security market data for models of dynamic economies, *Journal of Political Economy* 99: 225-262.
- Hansen, L. P. and K. J. Singleton (1982), Generalized instrumental variables estimation of nonlinear rational expectations models, *Econometrica* 50: 1269-1288.
- He, H. and D. M. Modest (1995), Market frictions and consumption-based asset pricing, *Journal of Political Economy* 103: 94-117.
- Heaton, J. C. (1995), An empirical investigation of asset pricing with temporally dependent preference specifications, *Econometrica* 63: 681-717.
- Heaton, J. C. and D. J. Lucas (1996), Evaluating the effects of incomplete markets on risk sharing and asset pricing, *Journal of Political Economy* 104: 443-487.
- Heaton, J., C. and D. J. Lucas (2000), Portfolio choice and asset prices: The importance of entrepreneurial risk, *Journal of Finance* 55: 1163-.
- Jacobs, K. (1999), Incomplete markets and security prices: Do asset-pricing puzzles result from aggregation problems? *Journal of Finance* 54: 123-163.
- Jacobs, K. and K. Q. Wang (2004), Idiosyncratic consumption risk and the cross-section of asset returns, *Journal of Finance* 59: 2211-2252.
- Jagannathan, R. and N. R. Kocherlakota (1996), Why should older people invest less in stocks than younger people? *Federal Bank of Minneapolis Quarterly Review* 20: 11-23.

- Jorion, P. and W. N. Goetzmann (1999), Global stock markets in the twentieth century, *Journal of Finance* 54: 953-980.
- Kocherlakota, N. R. (1996), The equity premium: it's still a puzzle, *Journal of Economic Literature* 34: 42-71.
- Korniotis, G. M. (2005), External habit, incomplete markets, and the cross-section of expected returns, Working paper, University of Notre Dame.
- Krebs, T. (2002), Testable implications of consumption-based asset pricing with incomplete markets, *Journal of Mathematical Economics* 40: 191-206.
- Li, Y. (2001), Expected returns and habit persistence, *Review of Financial Studies* 14: 861-899.
- Lucas, R., Jr. (1978), Asset prices in an exchange economy, *Econometrica* 46: 1429-1446.
- Lustig, H. (2004), The market price of aggregate risk and the wealth distribution, Working paper, UCLA.
- Lynch, A. W. (1996), Decision frequency and synchronization across agents: Implications for aggregate consumption and equity returns, *Journal of Finance* 51: 1479-1497.
- Mace, B. J. (1991), Full insurance in the presence of aggregate uncertainty, *Journal of Political Economy* 99: 928-956.
- Malloy, C. J., T. J. Moskowitz and A. Vissing-Jorgensen (2005), Long-run stockholder consumption risk and asset returns, Working paper, University of Chicago.
- Mankiw, N. G. (1986), The equity premium and the concentration of aggregate shocks, *Journal of Financial Economics* 17: 211-219.
- Mankiw, N. G. and M. D. Shapiro (1986), Risk and return: consumption beta versus market beta, *Review of Economics and Statistics* 68: 452-459.
- Mankiw, N. G. and S. P. Zeldes (1991), The consumption of stockholders and nonstockholders, *Journal of Financial Economics* 29: 97-112.
- Marshall, A. (1920), *Principles of Economics: An Introductory Volume* (Macmillan, London).
- McGrattan, E. R. and E. C. Prescott (2001), Taxes, regulations, and asset prices, Working paper, Federal Reserve Bank of Minneapolis.

- Mehra, R. (1998), On the volatility of stock prices: An exercise in quantitative theory, *International Journal of Systems Science* 29: 1203-1211.
- Mehra, R. and E. C. Prescott (1980), The equity premium: A puzzle, working paper, Carnegie-Mellon University.
- Mehra, R. and E. C. Prescott (1985), The equity premium: A puzzle, *Journal of Monetary Economics* 15: 145-161.
- Mehra, R. and E. C. Prescott (1988), The equity premium: A solution? *Journal of Monetary Economics* 22: 133-136.
- Mehra, R. and E. C. Prescott (2003), The equity premium in retrospect, in G. M. Constantinides, M. Harris, and R. Stulz, eds., *Financial Markets and Asset Pricing: Handbook of the Economics of Finance*, Volume 1B, Handbooks in Economics 21 (North-Holland, Amsterdam)
- Menzly, L., T. Santos and P. Veronesi (2004), Understanding predictability, *Journal of Political Economy* 112: 1-47.
- Merton, R. C. (1973), An intertemporal capital asset pricing model, *Econometrica* 41: 867-887.
- Negishi, T. (1960), Welfare economics and existence of an equilibrium for a competitive economy, *Metroeconomica* 12: 92-97.
- Parker, J. A. and C. Julliard (2005), Consumption risk and the cross section of expected stock returns, *Journal of Political Economy* 113: 185-222.
- Poterba, J. M. (2001), The rise of the “equity culture”: U.S. stockownership patterns, 1989-1998, Working paper, MIT.
- Rietz, T. A. (1988), The equity risk premium: a solution, *Journal of Monetary Economics* 22: 117-131.
- Rubinstein, M. (1976), The valuation of uncertain income streams and the pricing of options, *Bell Journal of Economics* 7: 407-425.
- Ryder, H. E. and G. M. Heal (1973), Optimum growth with intertemporally dependent preferences, *Review of Economic Studies* 40: 1-43.
- Semenov, A. (2005), High-order consumption moments and asset pricing, working paper, York University.

- Storesletten, K., C. I. Telmer and A. Yaron (2004), Consumption and risk sharing over the life cycle, *Journal of Monetary Economics*, 51: 609-633.
- Storesletten, K., C. I. Telmer and A. Yaron (2001), Asset pricing with idiosyncratic risk and overlapping generations, Working paper, Carnegie Mellon University.
- Sundaresan, S. M. (1989), Intertemporally dependent preferences and the volatility of consumption and wealth, *Review of Financial Studies* 2: 73-89.
- Vissing-Jorgensen, A. (2002), Limited asset market participation and the elasticity of intertemporal substitution, *Journal of Political Economy* 110: 825-853.
- Weil, P. (1989), The equity premium puzzle and the risk-free rate puzzle, *Journal of Monetary Economics* 24: 401-421.
- Weitzman, M. L. (2005), A unified Bayesian theory of equity puzzles, Working paper, Harvard University.