Rational Asset Prices

GEORGE M. CONSTANTINIDES*

ABSTRACT
The mean, covariability, and predictability of the return of different classes of financial assets challenge the rational economic model for an explanation. The unconditional mean aggregate equity premium is almost seven percent per year and remains high after adjusting downwards the sample mean premium by introducing prior beliefs about the stationarity of the price–dividend ratio and the (non)forecastability of the long-term dividend growth and price–dividend ratio. Recognition that idiosyncratic income shocks are uninsurable and concentrated in recessions contributes toward an explanation. Also borrowing constraints over the investors’ life cycle that shift the stock market risk to the saving middle-aged consumers contribute toward an explanation.

A central theme in finance and economics is the pursuit of a unified theory of the rate of return across different classes of financial assets. In particular, we are interested in the mean, covariability, and predictability of the return 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 past losing versus winning stocks.

The neoclassical rational economic model is a unified model that views these premia as the reward to risk-averse investors that process information rationally and have unambiguously defined preferences over consumption that typically (but not necessarily) belong to the von Neumann–Morgenstern class. Naturally, the theory allows for market incompleteness, market imperfections, informational asymmetries, and learning. The theory also allows for differences among assets for liquidity, transaction costs, tax status, and other institutional factors.

The cause of much anxiety over the last quarter of a century is evidence interpreted as failure of the rational economic paradigm to explain the price level and the rate of return of financial assets both at the macro and micro
levels. A celebrated example of such evidence, although by no means the only one, is the failure of the representative-agent rational economic paradigm to account for the large average premium of the aggregate return of stocks over short-term bonds and the small average return of short-term bonds from the last quarter of the 19th century to the present. Dubbed the "Equity Premium Puzzle" by Mehra and Prescott (1985), it has generated a cottage industry of rational and behavioral explanations of the level of asset prices and their rate of return.

Another example is the large increase in stock prices in the early and middle 1990s, which Federal Reserve Chairman Alan Greenspan decried as "Irrational Exuberance" even before the unprecedented further increase in stock prices and price-dividend ratios in the late 1990s.

My objective is to revisit some of this evidence and explore the extent to which the rational economic paradigm explains the price level and the rate of return of financial assets over the past 100+ years, both at the macro and micro levels.

In Section I, I reexamine the statistical evidence on the size of the unconditional mean of the aggregate equity return and premium. First, I draw a sharp distinction between conditional, short-term forecasts of the mean equity return and premium and estimates of the unconditional mean. I argue that the currently low conditional short-term forecasts of the return and premium do not lessen the burden on economic theory to explain the large unconditional mean equity return and premium, as measured by their sample average over the past 130 years. Second, I argue that even though one may introduce one's own strong prior beliefs and adjust downwards the sample-average estimate of the premium, the unconditional mean equity premium is at least 6 percent per year and the annual Sharpe ratio is at least 32 percent. These numbers are large and call for an economic explanation.

In Section II, I discuss limitations of the current theory to explain empirical regularities. I argue that per capita consumption growth covaries too little with the return of most classes of financial assets and this implies that the observed aggregate equity return, the long-term bond return, and the observed returns of various subclasses of financial assets are too large, too variable, and too predictable.

In the remaining sections, I revisit and examine the extent to which we can explain the asset returns by relaxing the assumptions of complete consumption insurance, perfect markets, and time-separable preferences. As the reader will readily observe—and I offer my apologies—my choice of issues is eclectic and mirrors in part my own research interests.

In Section III, I show that idiosyncratic income shocks concentrated in periods of economic recession play a key role in generating the mean equity premium, the low risk-free rate, and the predictability of returns. I argue that insufficient attention has been paid to the fact that the annual aggregate labor income exceeds annual dividends by a factor of over 20. Labor income is by far the single most important source of household savings and consumption. The shocks to labor income are uninsurable and persistent and arrive with greater frequency during economic contractions. Idiosyncratic
income shocks go a long way toward explaining the unconditional moments of asset returns and the predictability of returns. The construct of per capita consumption is largely irrelevant in explaining the behavior of asset returns because idiosyncratic income shocks are averaged out in per capita consumption.

In Section IV, I show that borrowing constraints over the life cycle play an important role in simultaneously addressing the above issues and the demand for bonds. I argue that insufficient attention has been paid to the consumers' life cycle consumption and savings decisions in a market with borrowing constraints. These considerations are important in addressing the limited participation of consumers in the capital markets, the irrelevance of the construct of per capita consumption, and the demand for short-term bonds by consumers with moderate risk aversion, given that equities earn on average a large premium over short-term bonds.

In Section V, I discuss the role of limited market participation. In Section VI, I discuss the role of habit persistence in addressing the same class of issues. In Section VII, I conclude that the observed asset returns do not support the case for abandoning the rational economic theory as our null hypothesis. Much more remains to be done to fully exploit the ramifications of the rational asset-pricing paradigm.

I. How Large Is the Equity Premium?

The average premium of the arithmetic rate of return of the S&P Composite Index over the risk-free rate, measured over the last 130 years, is almost 7 percent and the annual Sharpe ratio is 36 percent. If the equity premium is a stationary process, then the average premium is an unbiased estimate of the unconditional mean equity premium. One may introduce one's own prior beliefs and shave about 1 percent off the premium. The premium and the Sharpe ratio are still large and challenge economic theory for an explanation.

In Table I, I report the sample mean of the annual arithmetic aggregate equity return and of the equity premium. I proxy the aggregate equity return with the S&P Composite Index return. I proxy the annual risk-free rate with the rolled-over return on three-month Treasury bills and certificates. The reported real return is CPI-adjusted for inflation. Over the period 1872 to 2000, the sample mean of the real equity return is 8.9 percent and of the premium is 6.9 percent. Over the period 1926 to 2000, the sample mean of the equity return is 9.7 percent and that of the premium is 9.3 percent. Over the postwar period 1951 to 2000, the sample mean of the equity return is 9.9 percent and that of the premium is 8.7 percent. These sample means are large. Siegel (1998, 1999), Ibbotson Associates (2001), Ibbotson and Chen (2001), Dimson, Marsh, and Staunton (2002), Fama and French (2002), Mehra and Prescott (2002), and several others report the sample means of the equity return and premium in the United States and other countries and conclude that they are large. Some differences arise based on the proxy used for the risk-free rate.
I draw a sharp distinction between conditional, short-term forecasts of the mean equity return and premium and estimates of the unconditional mean. The conditional forecasts of the mean equity return and premium at the end of the 20th century and the beginning of the 21st are substantially lower than the estimates of the unconditional mean by at least three measures. First, based on evidence that price–dividend and price–earnings ratios forecast aggregate equity returns and that the values of these ratios prevailing at the beginning of the 21st century are well above their historic averages, Campbell and Shiller (1998) and Shiller (2000) forecast a conditional equity premium well below its sample average.1 Second, Claus and Thomas (2001)

1 Shiller (1984), Campbell and Shiller (1988a, 1988b), and Fama and French (1988) provide early evidence that the aggregate price–dividend and price–earnings ratios forecast aggregate equity returns. Goyal and Welch (1999) argue that the out-of-sample evidence is less convincing. I do not review here the debates and extensions relating to this literature. In the following paragraphs and in Appendix A, I argue that the forecastability results provide little, if any, guidance to my primary goal in this section, the estimation of the unconditional mean equity return.
calculate the expected aggregate equity premium to be a little above 3 percent in the period 1985 to 1998, based on analysts’ earnings forecasts. Third, Welch (2001) reports that the mean forecast among finance and economics professors for the one-year conditional equity premium is 3.5 percent in 2001, down from 6 percent in 1997. These findings are important in their own right and relevant in asset allocation.

However, the currently low conditional, short-term forecasts of the equity premium do not necessarily imply that the unconditional estimate of the mean premium is lower than the sample average. Therefore, the low conditional forecasts do not necessarily lessen the burden on economic theory to explain the large sample average of the equity return and premium over the past 130 years.

The predictability of aggregate equity returns by the price–dividend and price–earnings ratios raises the possibility that use of these financial ratios may improve upon the estimates of the unconditional mean equity return (and premium) that are based on the sample mean, an approach pursued earlier by Fama and French (2002). Over the period 1872 to 2000, the price–dividend ratio increased by a factor of 4.6 and the price–earnings ratio by a factor of 2.5. Over the period 1926 to 2000, the price–dividend ratio increased by a factor of 3.9 and the price–earnings ratio increased by a factor of 2.6. One may consider adjusting downwards the sample-mean estimate of the unconditional mean return on equity, but it is unclear by how much.

The size of the adjustment ought to relate to the perceived cause of the increase of these financial ratios. In the year 1998, 52 percent of the U.S. adult population held equity either directly or indirectly, compared to 36 percent of the adult population in 1989. This equitization has been brought about by the increased accessibility of information on the stock market, electronic trading, the growth of mutual funds, the growth of defined-contribution pension plans, and demographic changes. Other regime shifts include the advent of the technology/media/telecoms “new economy” and changes in the taxation of dividends and capital gains. Explanations of the price increase that rely on economic models that are less than fully rational include cultural and psychological factors and tap into the rich and burgeoning literature on behavioral economics and finance.

How does one process this information and adjust the sample mean estimate of the unconditional mean return and premium? To address this issue, I denote by \( v_t = \ln(P_t/X_t) \) the logarithm of the ratio of the price to the dividend.

---

2 The estimators employed in Fama and French (2002) and in this section are discussed in Appendix A.

3 The increase in these financial ratios should be interpreted with caution. The increase in the price–dividend ratio is due in part to an increase in share repurchases and a decrease in the fraction of dividend-paying firms.

4 I do not provide a systematic review of the offered explanations. Heaton and Lucas (1999), Shiller (2000), and McGrattan and Prescott (2001) provide lucid accounts of a number of these explanations in the context of both rational economic models and models that deviate from full rationality.
normalizing variable $X_t$, where the normalizing variable stands for the aggregate dividends, earnings, book equity, National Income, or some combination of these and other economic variables. I choose the normalizing variable $X_t$ in a way that I can plausibly assert that the log financial ratio is stationary. Over the sample period of length $T$ years, the mean annual (geometric) growth of the financial ratio $P_t/X_t$ is given by $(v_{T+1} - v_1)/T$. I define the adjusted estimator of the unconditional mean of the annual aggregate real equity return as the sample mean return, less some fraction beta of the sample mean annual growth of the financial ratio, $\hat{R}_{\text{SAMPLE}} - \beta(v_{T+1} - v_1)/T$. If the equity return and the log financial ratio are stationary processes, then the adjusted estimator is unbiased for any value of beta. However, the assumption of stationarity alone is insufficient to determine the value of beta.

The beta of the most efficient (mean squared error) adjusted estimator is equal to the slope coefficient of the regression of the sample mean return on the sample mean growth of the financial ratio, $(v_{T+1} - v_1)/T$. Since I have only one sample (of length $T$), I cannot run such a regression and must rely on information outside the sample and/or prior beliefs about the underlying economic model. In Appendix A, I present a set of sufficient conditions that imply that the beta of the most efficient estimator within this class of adjusted estimators is equal to one, when the adjustment is based on the price–dividend ratio. In addition to stationarity, the other main conditions are that the price–dividend ratio does not forecast the long-run growth in dividends and the long-run dividend growth does not forecast the price–dividend ratio. Adoption of the stationarity and (non)forecastability conditions requires strong prior beliefs.

In Table I, I report the mean annual growth of various financial ratios. Over the period 1951 to 2000, the mean annual growth of the price–dividend ratio is 3.4, the price–earnings ratio is 2.7, the price–book equity ratio is 3.2, and the price–National Income ratio is 1.3. Even if I subtract the entire mean annual growth of the price–earnings ratio from the sample mean, the adjusted estimate of the unconditional mean premium is 6.0 percent and is large. The corresponding estimate over the 1926 to 2000 period is 8.0 percent.

An alternative approach is to consider the longer sample period 1872 to 2000. Over this period, the mean annual growth of the price–dividend ratio and price–earnings ratio is 1.2 percent and 0.7 percent, respectively. Thus, this type of adjustment is largely a nonissue over the full sample. Essentially, the change in the financial ratios is “amortized” over 129 years and makes little difference in the estimate. Over the full period 1872 to 2000, the sample mean equity premium is 6.9 percent and the annual Sharpe ratio is

---

5 The ratio of the stock market value to the National Income is discussed in Mehra (1998).

6 A caveat is in order: Without additional assumptions, it is unclear what optimality properties (beyond unbiasedness) are associated with this class of estimators. Neither least squares, nor maximum likelihood, nor Bayesian methods motivate this class of estimators without further assumptions.
36 percent. Any adjustment with the average growth of the financial ratios still leaves the unconditional mean premium large and in need of an economic explanation.

II. Limitations of the Current Theory

The neoclassical rational-expectations economic model parsimoniously links the returns of all assets to the per capita consumption growth through the Euler equations of consumption (see Merton (1973), Rubinstein (1976), Lucas (1978), and Breeden (1979)). 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 covariability typically backfire, implying that the observed risk-free rate is too low and has too low variance. 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 the 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.7

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. The rejections tell us that 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

7 Grossman and Shiller (1981), Hansen and Singleton (1982), Ferson and Constantinides (1991), Hansen and Jagannathan (1991), and many others test and reject the Euler equations of consumption. Mehra and Prescott (1985) calibrate an economy to match the process of consumption growth. They demonstrate 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) stresses that the puzzle is a dual puzzle of the observed too high equity return and too low 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,\textsuperscript{8} modified probability distributions to admit rare but disastrous market-wide events,\textsuperscript{9} incomplete markets,\textsuperscript{10} and market imperfections.\textsuperscript{11} They also include a better understanding of data problems such as limited participation of consumers in the stock market,\textsuperscript{12} temporal aggregation,\textsuperscript{13} and the survival bias of the U.S. capital market.\textsuperscript{14} 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 excellent reviews in the textbooks by Campbell, Lo, and MacKinlay (1997) and Cochrane (2001), and in the articles by Cochrane and Hansen (1992), Kocherlakota (1996), Cochrane (1997), Campbell (2001, 2002), and Mehra and Prescott (2002).

III. Idiosyncratic Income Shocks and Incomplete Markets

A. The Role of 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 also accentuates its implications. Investors require a hefty equity premium in order to be induced to hold equities. In sum, this is the argument that I formalize below and address the predictability of asset returns and their unconditional moments.

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 aggregate (as opposed to disaggregate) consumption is,


\textsuperscript{9} The merits of this explanation are discussed in Mehra and Prescott (1988) and Rietz (1988).


\textsuperscript{13} Heaton (1995), Lynch (1996), and Gabaix and Laibson (2001).

\textsuperscript{14} See Brown, Goetzmann, and Ross (1995). However, Jorion and Goetzmann (1999, Table 6) find that the average real capital gain rate of a U.S. 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.
and how hard we should push aggregate or per capita consumption to explain returns. At a theoretical level, aggregate consumption is a meaningful economic construct if the market is complete or effectively so.\textsuperscript{15} In a complete market, heterogeneous households are able to equalize, state by state, their marginal rate of substitution. 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.\textsuperscript{16} 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), Mehra and Prescott (1985), and Mankiw (1986) suggest the potential of enriching the asset-pricing implications of the representative-household paradigm, by relaxing the assumption of complete markets.\textsuperscript{17} Constantinides and Duffie (1996) find 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 theory requires that the idiosyncratic income shocks must have three properties in order to explain the returns on financial assets. First, they must be \textit{uninsurable}. If the income shocks can be insured, then the household consumption growth is equal, state by state, to the aggregate consumption growth, and household consumption growth cannot do better than aggregate consumption growth in explaining the returns. Second, the income shocks must be \textit{persistent}. If the shocks are transient, then households can smooth their consumption by borrowing or by drawing down their savings.\textsuperscript{18} Third, the income shocks must be \textit{heteroscedastic}, with \textit{counter-cyclical} conditional variance.

A good example of a major uninsurable income shock is job loss. Job loss is \textit{uninsurable} because unemployment compensation is inadequate. Layoffs have \textit{persistent} implications on household income, even though the laid-off

\textsuperscript{15} The market is effectively complete when all households have preferences that imply one-fund or two-fund separation.

\textsuperscript{16} See Negishi (1960), Constantinides (1982), and Mehra and Prescott (1985, an unpublished earlier draft).

\textsuperscript{17} There is an extensive literature on the hypothesis of complete consumption insurance. See Cochrane (1991), Mace (1991), Altonji, Hayashi, and Kotlikoff (1992), and Attanasio and Davis (1997).

\textsuperscript{18} Aiyagari and Gertler (1991) and Heaton and Lucas (1996) find that consumers facing \textit{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.
workers typically find another job quickly. Layoffs are countercyclical as they are more likely to occur in recessions.

The first implication of the theory is an explanation of the countercyclical 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.

Since the proposition demonstrates the existence of equilibrium in frictionless markets, it implies that the Euler equations of household (but not necessarily of per capita) consumption must hold. Furthermore, since the given price processes have embedded in them whatever predictability of returns 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.

B. Empirical Evidence and Generalizations

Brav et al. (2002) provide empirical evidence of the importance of uninsurable idiosyncratic income risk on pricing. They estimate the RRA coefficient and test the set of Euler equations of household consumption on the premium of the value-weighted and the equally weighted market portfolio return over the risk-free rate, and on the premium of value stocks over growth stocks. They do not reject the Euler equations of household consumption with RRA coefficient between two and four, although they reject the Euler equations of per capita consumption with any value of the RRA coefficient. A RRA coefficient between two and four is economically plausible.

Open questions remain that warrant further investigation. According to the theory in Constantinides and Duffie (1996), periods with frequent and large uninsurable idiosyncratic income shocks are associated with both dispersed cross-sectional distribution of the household consumption growth and low stock returns. An interesting empirical question is which moments of the
cross-sectional distribution of the household consumption growth capture the dispersion. Brav et al. (2002) find that, in addition to the mean and variance, the skewness of the cross-sectional distribution is important in explaining the equity premium.

Krebs (2002) provides 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 extends the Constantinides and Duffie (1996) model that has only lognormal idiosyncratic income shocks 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 demonstrates 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 provides 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. In Appendix B, I present an example based on Krebs (2002).

A promising direction for future research is to address the relation between the equity return and the higher-order cross-sectional moments of household consumption with Monte Carlo methods. Another promising direction is to instrument the hard-to-observe time-series changes in the cross-sectional distribution with Labor Bureau statistics.

IV. The Life Cycle and Borrowing Constraints

A. Borrowing Constraints over the Life Cycle

Borrowing constraints provide an endogenous partial explanation for the limited participation of young consumers in the stock market. Constantinides et al. (2002a) construct an overlapping-generations exchange economy in which consumers live for three periods. In the first period, a period of human capital acquisition, the consumer receives a relatively low endowment income. In the second period, the consumer is employed and receives wage income subject to large uncertainty. In the third period, the consumer retires and consumes the assets accumulated in the second period. The 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. 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.

The model serves as a useful laboratory to address a range of economic issues. Campbell et al. (2001), and Constantinides, Donaldson, and Mehra (2001) address the cost of Social Security reform. Storesletten et al. (2001) explore the interaction of life-cycle effects and the uninsurable wage income shocks and find that the interaction plays an important role in explaining asset returns. Heaton and Lucas (1999) explore whether changes in market participation patterns account for the recent rise in stock prices and find that they do not.

B. Utility of Wealth—An Old Folks’ Tale

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 the asset returns, in the context of a representative-consumer economy with time separable preferences. Mankiw and Shapiro (1986) find 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 aggregate stock market return in the Euler equations of consumption.

I present an old folks’ tale, introduced in Constantinides, Donaldson, and Mehra (2002a, 2002b), that accomplishes this goal without introducing Epstein–Zin (1991) preferences or preferences defined directly over wealth.

---

21 See also the discussion in the related papers by Bodie, Merton, and Samuelson (1992), Jagannathan and Kocherlakota (1996), Bertaut and Haliassos (1997), Cocco, Gomes, and Maenhout (1999), and Storesletten et al. (2001).

22 Friend and Blume (1975) explain 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. Epstein and Zin (1991) introduce a recursive preference structure that emphasizes the timing of the resolution of uncertainty. 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) introduce 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.
Old folks who are rich enough to be nontrivial investors in the capital markets care about their wealth just as much as younger folks do, even though the state of their health and their medical expenses account for their consumption patterns better than fluctuations of their wealth do. This simple observation takes us a long way toward understanding why the stock market return does a better job than the growth of aggregate consumption does in explaining asset returns.

In the context of an overlapping-generations economy, the major investors in the market are the middle-aged households at the saving phase of their life cycle. These households save with the objective to maximize the utility of their “consumption” in their middle and old age. The insight here is that “consumption” of the old consists of two components, direct consumption, $c_D$; and the “joy of giving,” $c_B$, in the form of inter vivos gifts and post mortem bequests. Since the old households’ direct consumption is constrained by the state of their health, the correlation between the direct consumption of the old and the stock market return is low, a prediction that is borne out empirically. Therefore, the balance of the old households’ wealth, $c_B$, is a fortiori highly correlated with the stock market return. In terms of a utility function of consumption at the old age, $u(c_D) + v(c_B)$, that is separable over direct consumption and bequests, the model predicts an Euler equation of consumption with marginal utility at the old age given by $v'(c_B)$ and not by $u'(c_D)$, where $c_B$ is proxied by the stock market value.

This model remains to be tested. Nevertheless, it reinforces the general point that per capita consumption measures neither the total consumption of the marginal investor in the stock market nor that part of the marginal investor’s consumption that is unconstrained by health and medical considerations.

V. Limited Stock Market Participation

Limited stock market participation is another potential culprit in understanding why models of per capita consumption do a poor job in explaining returns. Whereas we understood all along that many households whose consumption is counted in the measure of per capita consumption do not hold stocks, it took a paper by Mankiw and Zeldes (1991) to point out that the emperor has no clothes. Even though 52 percent of the U.S. 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. Furthermore, wealthy entrepreneurs may be inframarginal in the stock market if their wealth is tied up in private equity.

23 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 (1999, 2000), Poterba (2001), and the collected essays in Guiso, Haliassos, and Jappelli (2001).
Mankiw and Zeldes (1991) calculate the per capita food consumption of a subset of households, designated as asset holders according to a criterion of asset holdings above some threshold. They find that the implied RRA coefficient decreases as the threshold is raised. Brav and Geczy (1995) confirm their result by using the nondurables and services per capita consumption, reconstructed from the Consumer Expenditure Survey (CEX) database. Attanasio et al. (2002), Brav et al. (2002), and Vissing-Jorgensen (2002) find 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 et al. point out that the statistical evidence is weak and the results are sensitive to experimental design.

Limited stock market participation is a fact of life and empirical tests of the Euler equations of consumption should account for it. However, my interpretation of the empirical results is that recognition of limited stock market participation alone is insufficient to explain the returns on assets. Essentially, the subset of households that are marginal in the stock market are still subject to uninsurable idiosyncratic income risk and we should take that into account also in attempting to explain asset returns.

VI. Habit Persistence

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. Building on earlier work by Ryder and Heal (1973) and Sundaresan (1989), I demonstrate in Constantinides (1990) that 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. Habit persistence lowers the intertemporal elasticity of substitution in consumption, given the risk aversion. The mean equity premium is equal to the covariance of consumption growth with equity returns, divided by this elasticity. Therefore, given the risk aversion, habit persistence lowers the elasticity and raises the mean equity premium.24

There are several interesting variations of the above class of preferences. Pollak (1970) discusses a model of external habit persistence in which the consumer does not take into account the effect of current consumption on future preferences. Abel (1990) and Campbell and Cochrane (1999) address...

---

24 Ferson and Constantinides (1991) test the special case of the linear habit model in which the habit depends only on the first lag of own consumption and report that the habit model performs better than the time-separable model and that the habit persistence parameter is economically and statistically significant. See also Hansen and Jagannathan (1991) and Heaton (1995).
the equity premium in the context of models with external habit persistence. In particular, the latter introduce a nonlinear specification of habit, reverse-engineered to keep the variability of the interest rate low. The large average equity premium, the predictability of long-horizon returns, and the behavior of equity prices along the business cycle are induced by a volatile RRA coefficient that has the value of 80 in the steady state and much higher still in economic recessions. Calibrated with the actual history of aggregate consumption, the model hits the aggregate price–dividend ratio in a number of periods but misses it in the 1950s and 1990s.

A promising direction for future research is to endogenize the currently ad hoc specification of the nonlinear habit. Another direction is to address the predictability of asset returns and their behavior along the business cycle in a model that benefits from the added flexibility of the nonlinear specification of habit but keeps risk aversion low and credible with the specification of habit to be internal.

Empirical tests of consumption-based models that incorporate habit persistence and aimed at explaining asset returns produce mixed results.\(^{25}\) It is hardly surprising that the results on both the habit and the external habit persistence models are mixed. The National Income and Product Accounts (NIPA) per capita consumption series is an imperfect proxy of the consumption of investors that are marginal in the capital markets, given the earlier-identified problems of incomplete consumption insurance, limited participation of households in the capital markets, borrowing constraints, and the exclusion of bequests from the definition of consumption. Both NIPA per capita consumption and consumption surplus over habit have low covariance with asset returns. Nonlinear refinements in the definition of habit do not remedy the problem of low covariance with asset returns. Habit persistence may well gain in empirical relevance in explaining asset returns, once we correctly measure the consumption of the unconstrained marginal investors in the capital markets.

Habit persistence is already gaining ground as an ingredient of economic models addressing a diverse set of economic problems beyond asset pricing, including the consumption–saving behavior and the home-equity puzzle. Habit persistence is a sensible property of preferences. It is also a property that allows for the separate specification of the RRA coefficient and the intertemporal elasticity of substitution within the class of von Neuman–Morgenstern preferences.

\(^{25}\) Ferson and Harvey (1992) report positive results for the linear external habit model. Wachter (2001) reports that long lags of consumption growth predict the short-term interest rate, as implied by the nonlinear external habit model. Li (2001) reports that in both the linear and the nonlinear external habit models, the surplus consumption over habit has limited success in explaining the time series of the premia of stock and bond portfolios. Menzly, Santos, and Veronesi (2001) develop an external habit model and report that it helps explain the cross section of asset returns.
VII. Concluding Remarks

I examine the observed asset returns and conclude that the evidence does not support the case for abandoning the rational economic model. I argue that the standard model is greatly enhanced by relaxing some of its assumptions. In particular, I argue 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 argue that 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 trace 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 argue 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 the price behavior along the business cycle with credibly low risk-aversion coefficient.

I believe that the integration of the notions of 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 within the class of rational asset-pricing models.

At the same time, I believe that specific deviations from rationality in the agents’ choices and in the agents’ processing of information potentially enhance the realism and economic analysis of certain phenomena on a case-by-case basis. However, several examples of apparent deviation from rationality may be reconciled with the rational economic paradigm, once we recognize that rational investors have incomplete knowledge of the fundamental structure of the economy and engage in learning. In any case, the collection of these deviations from rationality does not yet amount to a new economic paradigm that challenges the rational economic model.

It has been more than 60 years since Keynes (1936) wrote about animal spirits, and 15 since Shiller (1984) wrote about noise traders and DeBondt and Thaler (1985) wrote about stock market overreaction. I have yet to see an unambiguously articulated set of principles that emerges from the kalei-
doscope of these clinical investigations and that is put forth as an alterna-
tive to the rational economic paradigm. Serious scholars are keenly aware of
this criticism and hard at work to address it. Until such a paradigm is put
forth and is empirically vindicated, the rational economic paradigm remains
our principal guide to economic behavior.

Appendix A.

Estimation of the Unconditional Mean Return on Equity

I define the adjusted estimator of the unconditional mean of the annual
aggregate arithmetic real return on equity as

$$
\hat{R}_x = T^{-1} \sum_{t=1}^{T} R_{t+1} - \beta T^{-1}(v_{T+1} - v_1) = \hat{R}_{\text{SAMPLE}} - \beta T^{-1}(v_{T+1} - v_1). \quad (A1)
$$

The term $v_t = \ln(P_t/X_t)$ is the logarithm of the price of aggregate equity,
normalized with the variable $X_t$, where $X_t$ stands for the aggregate divi-
dends, earnings, book equity, National Income, or some other economic variable.

I assume that $R_t$ and $v_t$ are stationary processes. Then $E[v_{T+1} - v_1] = 0$
and $\hat{R}_x$ is an unbiased estimator of the unconditional mean equity return.

Note that the assumption of stationarity alone does not determine the value
of the parameter beta that provides the most efficient estimator of the un-
conditional mean equity return. The variance of the estimator $\hat{R}_x$ is

$$
\text{var}(\hat{R}_x) = \text{var}(\hat{R}_{\text{SAMPLE}}) - 2\beta \text{cov}(\hat{R}_{\text{SAMPLE}}, T^{-1}(v_{T+1} - v_1)) \\
+ \beta^2 \text{var}(T^{-1}(v_{T+1} - v_1)) \quad (A2)
$$

and is minimized when beta is set equal to

$$
\beta^* = \frac{\text{cov}(\hat{R}_{\text{SAMPLE}}, T^{-1}(v_{T+1} - v_1))}{\text{var}(T^{-1}(v_{T+1} - v_1))}. \quad (A3)
$$

The beta of the most efficient (mean squared error) estimator is equal to
the slope coefficient of the regression of $\hat{R}_{\text{SAMPLE}}$ on $T^{-1}(v_{T+1} - v_1)$.

Since I have only one sample of length $T$, I cannot run such a regression
and must rely on information outside the sample and/or prior beliefs about
the underlying economic model. Essentially, within the sample of length $T$, I
can examine the high-frequency behavior of the joint time series $R_t$ and $v_t$,
but I need to assert my prior beliefs on how these findings relate to the
behavior of the joint time series at the $T$-year frequency.

For example, consider the case in which $v_t$ stands for the log price–
dividend ratio. Since a high price–dividend ratio forecasts in-sample low
long-horizon returns, it is a plausible prior belief that it also forecasts low
$T$-horizon returns, $\text{cov}(\hat{R}_{\text{SAMPLE}}, v_1) < 0$, for $T = 50$ years (1951 to 2000) or
\( T = 129 \) years (1872 to 2000). It is also a plausible prior belief that periods of high returns are not followed by low price–dividend ratios, that is, it is plausible to believe that \( \text{cov}(\hat{R}_{\text{SAMPLE}}, v_{T+1}) \geq 0 \). Then equation (A3) implies that the beta of the most efficient estimator is positive.

I present a set of sufficient (but not necessary) conditions that imply that the beta of the most efficient estimator in the class \( \hat{R}_x \) equals one. Let \( v_t \) stand for the log price–dividend ratio and assume the following: (1) the returns and the price–dividend ratio are stationary, (2) the price–dividend ratio does not forecast the growth in dividends, (3) dividend growth does not forecast the price–dividend ratio, (4) the price–dividend ratio does not forecast the difference in the conditional variance of the capital gain rate and the dividend growth rate, and (5) the difference in the conditional variance of the capital gain rate and the dividend growth rate does not forecast the price–dividend ratio. To prove the claim, I use a Taylor-series expansion:

\[
\Delta v_{t+1} = \Delta P_{t+1}/P_t - \Delta D_{t+1}/D_t - k_{t+1} \tag{A4}
\]

where

\[
k_{t+1} = (\Delta P_{t+1}/P_t)^2/2 - (\Delta D_{t+1}/D_t)^2/2
\]

and write the sample mean of the arithmetic return as

\[
\hat{R}_{\text{SAMPLE}} = T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta P_{t+1}/P_t\} = T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta D_{t+1}/D_t + k_{t+1} + \Delta v_{t+1}\} = T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta D_{t+1}/D_t + k_{t+1}\} + T^{-1}(v_{T+1} - v_1). \tag{A5}
\]

I substitute the value of \( \hat{R}_{\text{SAMPLE}} \) from equation (A5) into equation (A3) and obtain the result that the variance of the estimator is minimized when the value of beta is one:

\[
\beta^* = \frac{\text{cov}\left(\sum_{t=1}^{T} \Delta D_{t+1}/D_t, (v_{T+1} - v_1)\right)}{\text{var}(v_{T+1} - v_1)} + \frac{\text{cov}\left(\sum_{t=1}^{T} \Delta P_{t+1}/P_t, v_{T+1}\right)}{\text{var}(v_{T+1} - v_1)} - \frac{\text{cov}\left(\sum_{t=1}^{T} \Delta D_{t+1}/D_t, v_{T+1}\right)}{\text{var}(v_{T+1} - v_1)} + \frac{\text{cov}\left(\sum_{t=1}^{T} k_{t+1}, (v_{T+1} - v_1)\right)}{\text{var}(v_{T+1} - v_1)} + 1 \tag{A6}
\]

\[= 1. \]
The first term in equation (A6) is approximately zero because the stationarity of the price–dividend ratio implies

$$\text{cov} \left( \sum_{t=1}^{T} D_{t+1}/P_t, v_{T+1} \right) \approx \text{cov} \left( \sum_{t=1}^{T} D_{t+1}/P_t, v_1 \right).$$

(A7)

The second term in equation (A6) is zero because, by assumption, the dividend growth rate does not forecast the price–dividend ratio. The third term is zero because, by assumption, the price–dividend ratio does not forecast the dividend growth. Finally, the fourth term is zero because, by assumption, the price–dividend ratio does not forecast and is not forecasted by the difference of the conditional variance of the capital gain rate and the dividend growth rate.

Thus, when $X_t$ stands for the dividends and conditions (1)–(5) hold, the minimum variance estimator in the class of estimators given by equation (A1) is

$$\hat{R}_D = \hat{R}_{\text{SAMPLE}} - T^{-1} (v_{T+1} - v_1)$$

$$= T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta D_{t+1}/D_t\} + T^{-1} \sum_{t=1}^{T} k_{t+1}.$$ 

(A8)

Fama and French (2002) report adjusted estimates of the unconditional mean return (and premium) based on the fundamentals dividends and earnings. Specifically, their estimate of the expected stock return based on the dividend growth model is equivalent to $T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta D_{t+1}/D_t\}$ and their biased-adjusted estimate is equivalent to $T^{-1} \sum_{t=1}^{T} \{D_{t+1}/P_t + \Delta D_{t+1}/D_t\} + T^{-1} \sum_{t=1}^{T} k_{t+1}$. Ibbotson and Chen (2001) also report adjusted estimates of the unconditional mean return (and premium) based on dividends, income, earnings, payout ratio, book equity, and National Income.

**Appendix B.**

**Extension of the Constantinides and Duffie (1996) Model**

I illustrate an extension of the Constantinides and Duffie (1996) model along the lines of Krebs (2002). The extension provides theoretical justification as to why it may be hard to detect empirically in the low-order cross-sectional moments of household consumption growth the rare but catastrophic shocks that play a major role in driving asset prices.

The $i$th household’s consumption, $c_{i,t}$, follows the process

$$\frac{c_{i,t}}{c_{i,t-1}} = \frac{c_t}{c_{t-1}} X_{i,t} \eta_{i,t}.$$ 

(B1)
The random variables \( \{ \eta_{i,t} \} \) have the following properties: Distinct subsets of \( \{ \eta_{i,t} \} \) are independent; for all \( i \) and \( t \); \( \eta_{i,t} \) is independent of \( c_{i-1}, c_t, c_{i,t-1}, X_{i,t} \), and the asset prices; and \( E[\eta_{i,t}] = 1 \). Since the random variables \( \{ \eta_{i,t} \} \) are independent of the asset prices, they do not contribute to the equity premium. One may choose to view them as observation error, but does not have to.

In the Constantinides and Duffie (1996) model, the idiosyncratic income shocks are lognormal: \( X_{i,t} = e^{b_i \varepsilon_{i,t}} \) with \( \varepsilon_{i,t} \) normal and \( \eta_{i,t} = 1 \). The conditional variance, \( b_i^2 \), explains the risk premia because it is modeled as countercyclical and correlated with the stock returns. Whereas Brav et al. (2002) find that the pricing kernel \( I^{-1} \sum_{i=1}^t (c_{i,t}/c_{i,t-1})^{-\alpha} \) goes a long way toward explaining the equity premium and the value-versus-growth premium, they also find little evidence that the conditional variance, \( b_i^2 \), is correlated with stock returns, or indeed whether the time series of this variance has any discernible pattern relative to the business cycle. I build this feature in the model by choosing a binomial distribution for \( X_{i,t} \).

I assume that the random variables \( \{ X_{i,t} \} \) have the following properties: Distinct subsets of \( \{ X_{i,t} \} \) are independent; for all \( i \) and \( t \), \( X_{i,t} \) is independent of \( c_{i-1}, c_t, c_{i,t-1} \) and \( X_{i,t-1} \); and \( X_{i,t} \) has the following binomial distribution:

\[
X_{i,t} = \begin{cases} 
1 - y_t^{1-\alpha} \pi^{1+a^{-1}} & \text{with probability } 1 - \pi \\
y_t^{-\alpha} \pi^{a^{-1}} & \text{with probability } \pi,
\end{cases}
\]

where \( 0 < \pi < 1 \), and \( \alpha \) is the constant RRA coefficient. The variable \( y_t, y_t > 0 \) is defined shortly. Since

\[
E \left[ \frac{c_{i,t}}{c_{i,t-1}} \mid y_t, c_t, c_{i,t-1} \right] = \frac{c_t}{c_{t-1}}, \quad (B2)
\]

arguments along the lines in Constantinides and Duffie (1996) identify \( c_t \) as the per capita consumption.

The time-\( t \) expectation of the \( i \)th household’s marginal rate of substitution, conditional on \( \{ c_t/c_{i-1}, y_t \} \), is

\[
E \left[ e^{-\rho} \left( \frac{c_{i,t}}{c_{i,t-1}} \right)^{-\alpha} \mid c_t, c_{i,t-1}, y_t \right] \\
= e^{-\rho} \left( \frac{c_t}{c_{t-1}} \right)^{-\alpha} \left( (1 - \pi)^{1+a} (1 - y_t^{-\alpha} \pi^{1+a^{-1}})^{-\alpha} + y_t \right) E[\eta_{i,t}^{-\alpha}] \\
\approx e^{-\rho} \left( \frac{c_t}{c_{t-1}} \right)^{-\alpha} (1 + y_t) E[\eta_{i,t}^{-\alpha}], \quad \text{for } \pi \ll 1. \quad (B4)
\]
I define the variable $y_t$ implicitly with the equation

$$e^{-\rho \left( \frac{c_t}{c_{t-1}} \right)^{-\alpha}} (1 + y_t) E[\eta_{t+1}^{\alpha}] = M_t,$$

where $M_t$ is the pricing kernel that supports the given joint process of aggregate income, asset prices, and dividends. By construction, it follows that any individual household’s marginal rate of substitution, $e^{-\rho (c_t/c_{t-1})^{-\alpha}}$, supports the given joint process of aggregate income, asset prices, and dividends.

Finally, I demonstrate that the variance, skewness, and higher moments of the cross-sectional distribution of the households’ consumption growth need not bear any relationship to asset returns and the business cycle. This is despite the fact that each individual household’s marginal rate of substitution supports the given joint process of aggregate income, asset prices, and dividends.

The $N$th central moment, $N \geq 1$, of the households’ logarithmic consumption growth is the sum of the $N$th central moments of $\ln(c_t/c_{t-1})$, $\ln(X_{i,t})$, and $\ln(\eta_{i,t})$, given the assumed independence of $c_t/c_{t-1}, X_{i,t},$ and $\eta_{i,t}$. It is easily shown that

$$\lim_{\pi \to 0} E[(\ln X_{i,t})^N] = 0, \quad N \geq 1.$$

If the probability of the idiosyncratic consumption shocks is sufficiently low, $\pi \ll 1$, the central moments of the households’ consumption growth are driven by the corresponding central moments of the per capita consumption growth and $\eta_{i,t}$. These moments need not bear any pattern relating to the business cycle and need not be correlated in any particular way with the asset returns. Despite this, each individual household’s marginal rate of substitution supports the given joint process of aggregate income, asset prices, and dividends. The illustration explains why it may be empirically difficult or infeasible to detect the idiosyncratic consumption shocks in the cross-sectional moments of household consumption growth.

REFERENCES


Ait-Sahalia, Yacine, Jonathan A. Parker, and Motohiro Yogo, 2001, Luxury goods and the equity premium, Working paper 8417, NBER.


Goyal, Amit, and Ivo Welch, 1999, Predicting the equity premium, Working paper, UCLA.


Negishi, Takashi, 1960, Welfare economics and existence of an equilibrium for a competitive economy, Metroeconomica 12, 92–97.


