WILEY-BLACKWELL

American Finance Association

Asset Pricing at the Millennium
Author(s): John Y. Campbell
Source: *The Journal of Finance*, Vol. 55, No. 4, Papers and Proceedings of the Sixtieth Annual Meeting of the American Finance Association, Boston, Massachusetts, January 7–9, 2000 (Aug., 2000), pp. 1515–1567
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/222372
Accessed: 27/09/2009 13:06

Asset Pricing at the Millennium

JOHN Y. CAMPBELL*

ABSTRACT

This paper surveys the field of asset pricing. The emphasis is on the interplay between theory and empirical work and on the trade-off between risk and return. Modern research seeks to understand the behavior of the stochastic discount factor (SDF) that prices all assets in the economy. The behavior of the term structure of real interest rates restricts the conditional mean of the SDF, whereas patterns of risk premia restrict its conditional volatility and factor structure. Stylized facts about interest rates, aggregate stock prices, and cross-sectional patterns in stock returns have stimulated new research on optimal portfolio choice, intertemporal equilibrium models, and behavioral finance.

This paper surveys the field of asset pricing. The emphasis is on the interplay between theory and empirical work. Theorists develop models with testable predictions; empirical researchers document "puzzles"—stylized facts that fail to fit established theories—and this stimulates the development of new theories.

Such a process is part of the normal development of any science. Asset pricing, like the rest of economics, faces the special challenge that data are generated naturally rather than experimentally, and so researchers cannot control the quantity of data or the random shocks that affect the data. A particularly interesting characteristic of the asset pricing field is that these random shocks are also the subject matter of the theory. As Campbell, Lo, and MacKinlay (1997, Chap. 1, p. 3) put it:

What distinguishes financial economics is the central role that uncertainty plays in both financial theory and its empirical implementation. The starting point for every financial model is the uncertainty facing investors, and the substance of every financial model involves the impact of uncertainty on the behavior of investors and, ultimately, on mar-

ket prices. . . . The random fluctuations that require the use of statistical theory to estimate and test financial models are intimately related to the uncertainty on which those models are based.

For roughly the last 20 years, theoretical and empirical developments in asset pricing have taken place within a well-established paradigm. This paradigm emphasizes the structure placed on financial asset returns by the assumption that asset markets do not permit the presence of arbitrage opportunities—loosely, opportunities to make riskless profits on an arbitrarily large scale. In the absence of arbitrage opportunities, there exists a "stochastic discount factor" that relates payoffs to market prices for all assets in the economy. This can be understood as an application of the Arrow–Debreu model of general equilibrium to financial markets. A state price exists for each state of nature at each date, and the market price of any financial asset is just the sum of its possible future payoffs, weighted by the appropriate state prices. Further assumptions about the structure of the economy produce further results. For example, if markets are complete then the stochastic discount factor is unique. If the stochastic discount factor is linearly related to a set of common shocks, then asset returns can be described by a linear factor model. If the economy has a representative agent with a well-defined utility function, then the SDF is related to the marginal utility of aggregate consumption. Even recent developments in behavioral finance, which emphasize nonstandard preferences or irrational expectations, can be understood within this paradigm.

From a theoretical perspective, the stability of the paradigm may seem to indicate stagnation of the field. Indeed Duffie (1992, Pref., pp. xiii–xiv) disparages recent progress by contrasting it with earlier theoretical achievements:

> To someone who came out of graduate school in the mid-eighties, the decade spanning roughly 1969–79 seems like a golden age of dynamic asset pricing theory. . . . The decade or so since 1979 has, with relatively few exceptions, been a mopping-up operation.

Without denying the extraordinary accomplishments of the earlier period, I hope to show in this paper that the period 1979 to 1999 has also been a highly productive one. Precisely because the conditions for the existence of a stochastic discount factor are so general, they place almost no restrictions on financial data. The challenge now is to understand the economic forces that determine the stochastic discount factor or, put another way, the rewards that investors demand for bearing particular risks. We know a great deal more about this subject today than we did 20 years ago. Yet our understanding is far from perfect, and many exciting research opportunities remain.

Any attempt to survey such a large and active field must necessarily be limited in many respects. This paper concentrates on the trade-off between risk and return. Most of the literature on this subject makes the simplifying

assumption that investors have homogeneous information. I therefore neglect the theory of asymmetric information and its applications to corporate finance, market microstructure, and financial intermediation. I concentrate on the U.S. financial markets and do not discuss international finance. I mention issues in financial econometrics only in the context of applications to risk-return models, and I do not review the econometric literature on changing volatility and nonnormality of asset returns. I do not draw implications of the asset pricing literature for asset management, performance evaluation, or capital budgeting. I leave most continuous-time research, and its applications to derivative securities and corporate bonds, to the complementary survey of Sundaresan (2000). I draw heavily on earlier exposition in Campbell et al. (1997) and Campbell (1999). The latter paper reports comparative empirical results on aggregate stock and bond returns in other developed financial markets.

I. Asset Returns and the Stochastic Discount Factor

The basic equation of asset pricing can be written as follows:

$$P_{it} = \mathrm{E}_t[M_{t+1}X_{i,t+1}], \tag{1}$$

where P_{it} is the price of an asset i at time t ("today"), E_t is the conditional expectations operator conditioning on today's information, $X_{i,t+1}$ is the random payoff on asset i at time $t + 1$ ("tomorrow"), and M_{t+1} is the stochastic discount factor, or SDF. The SDF is a random variable whose realizations are always positive. It generalizes the familiar notion of a discount factor to a world of uncertainty; if there is no uncertainty, or if investors are risk-neutral, the SDF is just a constant that converts expected payoffs tomorrow into value today.

Equation (1) can be understood in two ways. First, in a discrete-state setting, the asset price can be written as a state-price-weighted average of the payoffs in each state of nature. Equivalently, it can be written as a probability-weighted average of the payoffs, multiplied by the ratio of state price to probability for each state. The conditional expectation in equation (1) is just that probability-weighted average. The absence of arbitrage opportunities ensures that a set of positive state prices exists and hence that a positive SDF exists. If markets are complete, then state prices and the SDF are unique.

Second, consider the optimization problem of an agent k with time-separable utility function $U(C_{kt}) + \delta U(C_{k,t+1})$. If the agent is able to freely trade asset i, then the first-order condition is

$$U'(C_{kt})P_{it} = \delta\mathrm{E}_t[U'(C_{k,t+1})X_{i,t+1}], \tag{2}$$

which equates the marginal cost of an extra unit of asset i, purchased today, to the expected marginal benefit of the extra payoff received tomorrow. Equation (2) is consistent with equation (1) for $M_{t+1} = \delta U'(C_{k,t+1})/U'(C_{kt})$, the discounted ratio of marginal utility tomorrow to marginal utility today. This marginal utility ratio, for investors who are able to trade freely in a set of assets, can always be used as the SDF for that set of assets.

Equation (1) allows for the existence of assets—or investment strategies—with zero cost today. If P_{it} is nonzero, however, one can divide through by P_{it} (which is known at time t and thus can be passed through the conditional expectations operator) to obtain

$$1 = E_t[M_{t+1}(1 + R_{i,t+1})], \tag{3}$$

where $(1 + R_{i,t+1}) \equiv X_{i,t+1}/P_{it}$. This form is more commonly used in empirical work.

The origins of this representation for asset prices lie in the Arrow–Debreu model of general equilibrium and in the application of that model to option pricing by Cox and Ross (1976) and Ross (1978), along with the Arbitrage Pricing Theory of Ross (1976). A definitive theoretical treatment in continuous time is provided by Harrison and Kreps (1979), and the discrete-time representation was first presented and applied empirically by Grossman and Shiller (1981). Hansen and Richard (1987) develop the discrete-time approach further, emphasizing the distinction between conditional and unconditional expectations. Textbook treatments are given by Ingersoll (1987) and Duffie (1992). Cochrane (1999) restates the whole of asset pricing theory within this framework.

How are these equations used in empirical work? A first possibility is to impose minimal theoretical structure, using data on asset returns alone and drawing implications for the SDF. Work in this style, including research that simply documents stylized facts about means, variances, and predictability of asset returns, is reviewed later in this section. A second possibility is to build a time-series model of the SDF that fits data on both asset payoffs and prices; work along these lines, including research on the term structure of interest rates, is reviewed in Section II. A third approach is microeconomic. One can use equation (2), along with assumed preferences for an investor, the investor's intertemporal budget constraint, and a process for asset returns or equivalently for the SDF, to find the investor's optimal consumption and portfolio rules. This very active area of recent research is reviewed in Section III. Fourth, one can assume that equation (2) applies to a representative investor who consumes aggregate consumption; in this case C_{kt} is replaced by aggregate consumption C_t, and equation (2) restricts asset prices in relation to consumption data. Work along these lines is discussed in Section IV. Finally, one can try to explain equilibrium asset prices as arising from the interactions of heterogeneous agents. Section V discusses models

with rational agents who are heterogeneous in their information, income, preferences, or constraints. Section VI discusses recent research on behavioral finance, in which some agents are assumed to have nonstandard preferences or irrational expectations, whereas other agents have standard preferences and rational expectations. Section VII concludes.

A. Mean and Variance of the SDF

A.1. The Real Interest Rate

Asset return data restrict the moments of the SDF. The one-period real interest rate is closely related to the conditional mean of the SDF, conditioning on information available at the start of the period. If there is a short-term riskless real asset f with a payoff of one tomorrow, then equation (1) implies that

$$\mathrm{E}_t M_{t+1} = P_{ft} = \frac{1}{1 + R_{f,t+1}}. \tag{4}$$

The expected SDF is just the real price of the short-term riskless real asset or, equivalently, the reciprocal of its gross yield.

Of course, there is no truly riskless one-period real asset in the economy. Short-term Treasury bills are riskless in nominal terms rather than real terms, and even inflation-indexed bonds have an indexation lag that deprives them of protection against short-term inflation shocks. In practice, however, short-term inflation risk is sufficiently modest in the United States and other developed economies that nominal Treasury bill returns are a good proxy for a riskless one-period real asset. This means that the conditional expectation of the SDF is pinned down by the expected real return on Treasury bills. This return is fairly low on average (Campbell (1999) reports a mean log return of 0.8 percent per year in quarterly U.S. data over the period 1947.2 to 1996.4). It is also fairly stable (the standard deviation is 1.76 percent in the same data set, and perhaps half of this is due to ex post inflation shocks). In fact, Fama (1975) argued that in the 1950s and 1960s the real interest rate was actually constant. Since the early 1970s, however, there have been some lower-frequency variations in the real interest rate; it was very low or even negative during the late 1970s, much higher in the early 1980s, and drifted lower during the late 1980s. A reasonable model of the SDF must therefore have a conditional expectation that is slightly less than one, does not move dramatically in the short run, but has some longer-term variation. Such behavior can be captured using persistent linear time-series models (standard in the literature on the term structure of interest rates, discussed in Section II.B.2) or regime-switching models (Gray (1996), Garcia and Perron (1996)).

There is also a great deal of research on the time-series properties of the nominal interest rate. One can define a "nominal stochastic discount factor" as the real SDF times the price level today, divided by the price level tomorrow. The expectation of the nominal SDF is the price of a short-term risk-free nominal asset. This work is less relevant for the equilibrium issues discussed in this paper, however, as the nominal SDF cannot be related to optimal consumption in the same way as the real SDF. Much of this literature is surveyed by Campbell et al. (1997, Chap. 11) and Sundaresan (2000).

A.2. Risk Premia

Risk premia restrict the volatility of the SDF. Comparing equation (3) for a risky asset and for the riskless asset, we have $0 = \mathrm{E}_t[M_{t+1}(R_{i,t+1} - R_{f,t+1})] = \mathrm{E}_t M_{t+1} \mathrm{E}_t(R_{i,t+1} - R_{f,t+1}) + \mathrm{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1})$. Rearranging, the expected excess return on any asset satisfies

$$\mathrm{E}_t(R_{i,t+1} - R_{f,t+1}) = \frac{-\mathrm{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1})}{\mathrm{E}_t M_{t+1}}. \tag{5}$$

The expected excess return is determined by risk, as measured by the negative of covariance with the SDF, divided by the expected SDF or equivalently the price of a riskless asset. An asset whose covariance with the SDF is large and negative tends to have low returns when the SDF is high, that is, when marginal utility is high. In equilibrium such an asset must have a high excess return to compensate for its tendency to do poorly in states of the world where wealth is particularly valuable to investors.

Because the correlation between the SDF and the excess return must be greater than minus one, the negative covariance in equation (5) must be less than the product of the standard deviations of the excess return and the SDF. Rearranging, we have

$$\frac{\sigma_t(M_{t+1})}{\mathrm{E}_t M_{t+1}} \geq \frac{\mathrm{E}_t(R_{i,t+1} - R_{f,t+1})}{\sigma_t(R_{i,t+1} - R_{f,t+1})}. \tag{6}$$

In words, the Sharpe ratio for asset i—the asset's risk premium divided by its standard deviation—puts a lower bound on the volatility of the SDF. The tightest lower bound is achieved by finding the risky asset, or portfolio of assets, with the highest Sharpe ratio. This bound was first stated by Shiller (1982). Hansen and Jagannathan (1991) have extended it to a setting with many risky assets but no riskless asset, showing how to construct a frontier relating the lower bound on the volatility of the SDF to the mean of the SDF. This frontier contains the same information as the familiar mean-variance efficient frontier relating the lower bound on the variance of a portfolio return to the mean portfolio return. The lower bound is achieved by an SDF that is a linear combination of a hypothetical riskless asset and the risky

assets under consideration. Hansen and Jagannathan also derive a tighter bound by using the restriction that the SDF must always be positive. Co-chrane and Hansen (1992) present further empirical results, and Hansen and Jagannathan (1997) extend the methodology to consider the pricing er-rors that can be made by a false economic model for the SDF. They show that the largest possible pricing errors are bounded by the standard deviation of the difference between the false SDF and the true SDF, in a manner analo-gous to equation (6).

As written, all the quantities in equation (6) are conditional on informa-tion at time t, that is, they have time subscripts. Fortunately it is simple to derive an unconditional version by returning to equation (3) and taking un-conditional expectations; the form of equation (6) is unchanged. The only subtlety is that the unconditional mean SDF is the unconditional mean price of a riskless asset, which is not the same as the reciprocal of the uncondi-tional mean riskless real interest rate.

A.3. The Equity Premium Puzzle

It is not hard to find assets that imply surprisingly large numbers for the volatility of the SDF. The aggregate U.S. stock market is the best-known example. In postwar quarterly U.S. data summarized by Campbell (1999) the annualized Sharpe ratio for a value-weighted stock index is about one-half, implying a minimum annualized standard deviation of 50 percent for the SDF. This is a large value for a random variable whose mean must be close to one and whose lower bound is zero. As we shall see later in the paper, it is also very large relative to the predictions of simple equilibrium models. The annualized standard deviation of aggregate consumption growth in postwar U.S. data is about one percent. A representative-agent model with power utility must therefore have a very large coefficient of relative risk aversion, on the order of 50, to match the standard deviation of the SDF. Mehra and Prescott (1985) first drew the attention of the profession to this phenomenon and named it the "equity premium puzzle."

Of course, there is considerable uncertainty about the moments that enter equation (6). Cecchetti, Lam, and Mark (1994) and Hansen, Heaton, and Luttmer (1995) develop statistical methods, based on Hansen's (1982) Gen-eralized Method of Moments (GMM), to estimate a confidence interval for the volatility of the SDF. Mean asset returns are particularly hard to esti-mate because, as Merton (1980) pointed out, the precision of the estimate depends on the total length of calendar time rather than the number of observations per se. Fortunately U.S. stock market data are available for a period of almost two centuries (Schwert (1990) presents data starting in 1802); this long span of data means that even a lower confidence bound on the volatility of the SDF is quite large.

Some authors have argued that these results are misleading. If academic studies focus on long-term U.S. data precisely because the economy and the stock market have performed so well, then there is an upward selection bias

in measured average U.S. returns (Brown, Goetzmann, and Ross (1995)). Most other developed stock markets have offered comparable returns to the United States in the postwar period (Campbell (1999)), but Jorion and Goetzmann (1999) show that price returns in many of these other markets were low in the early twentieth century; this may indicate the importance of selection bias, although it is possible that lower returns were compensated by higher dividend yields in that period.

Rietz (1988) argues that the U.S. data are misleading for a different reason. Investors may have rationally anticipated the possibility of a catastrophic event that has not yet occurred. This "peso problem" implies that sample volatility understates the true risk of equity investment. One difficulty with this argument is that it requires not only a potential catastrophe but one that affects stock market investors more seriously than investors in short-term debt instruments. Many countries that have experienced political upheaval or defeat in war have seen very low returns on short-term government debt and on equities. A peso problem that affects both asset returns equally will not necessarily affect the estimated volatility of the SDF. The major example of a disaster for stockholders that spared bondholders is the Great Depression of the 1930s, but of course this event is already included in long-term U.S. data.

A.4. Predictability of Aggregate Stock Returns

Further interesting results are available if one uses conditioning information. There is an enormous literature documenting the predictability of aggregate stock returns from past information, including lagged returns (Fama and French (1988a), Poterba and Summers (1988)), the dividend-to-price ratio (Campbell and Shiller (1988a), Fama and French (1988b), Hodrick (1992)), the earnings-to-price ratio (Campbell and Shiller (1988b)), the book-to-market ratio (Lewellen (1999)), the dividend payout ratio (Lamont (1998)), the share of equity in new finance (Nelson (1999), Baker and Wurgler (2000)), yield spreads between long-term and short-term interest rates and between low- and high-quality bond yields (Campbell (1987), Fama and French (1989), Keim and Stambaugh (1986)), recent changes in short-term interest rates (Campbell (1987), Hodrick (1992)), and the level of consumption relative to income and wealth (Lettau and Ludvigson (1999a)). Many of these variables are related to the stage of the business cycle and predict countercyclical variation in stock returns (Fama and French (1989), Lettau and Ludvigson (1999a)).

A number of econometric pitfalls are relevant for evaluating these effects. First, return predictability appears more striking at long horizons than at short horizons; the explanatory power of a regression of stock returns on the log dividend-to-price ratio, for example, increases from around two percent at a monthly frequency to 18 percent at an annual frequency and 34 percent at a two-year frequency in postwar quarterly U.S. data (Campbell (1999)). The difficulty is that the number of nonoverlapping observations decreases with the forecast horizon, and it is essential to adjust statistical inference for this. Standard adjustments work poorly when the size of the overlap is

large relative to the sample size; Richardson and Stock (1989) suggest an alternative approach to handle this case. Second, many of the variables that appear to predict returns are highly persistent, and their innovations are correlated with return innovations. Even when returns are measured at short horizons, this can lead to small-sample biases in standard test statistics. Nelson and Kim (1993) and others use Monte Carlo methods to adjust for this problem.

Despite these difficulties, the evidence for predictability survives at reasonable if not overwhelming levels of statistical significance. Most financial economists appear to have accepted that aggregate returns do contain an important predictable component. Even the recent increase in U.S. stock prices, which has weakened the purely statistical evidence for mean-reversion and countercyclical predictability, has not broken this consensus because it is difficult to rationalize the runup in prices with reasonable dividend or earnings forecasts and constant discount rates (Heaton and Lucas (1999)).

Conditioning information can be used to learn about the SDF in two different ways. First, one can create a "managed portfolio" that increases the portfolio weight on stocks when one or more predictive variables suggest that stock returns will be high. The managed portfolio can then be included in the basic Hansen–Jagannathan analysis. To the extent that the managed portfolio has a higher Sharpe ratio than the unmanaged stock index, the Hansen–Jagannathan volatility bound will be sharpened.

Second, one can explicitly track the time variation in expected returns and volatilities. Campbell (1987), Harvey (1989, 1991), and Glosten, Jagannathan, and Runkle (1993) use GMM techniques to do this. They find that some variables that predict returns also predict movements in volatility, but there is also substantial countercyclical variation in Sharpe ratios. These results could be used to construct a time-varying volatility bound for the SDF.

For future reference, I note that much empirical work uses logarithmic versions of the SDF equations reviewed in this section. If one assumes that the SDF and asset returns are conditionally jointly lognormal, then one can use the formula for the conditional expectation of a lognormal random variable Z, $\ln(\mathrm{E}_t[Z]) = \mathrm{E}_t(\ln(Z)) + 1/2\mathrm{Var}_t(\ln(Z))$. Applied to equation (4), this delivers an expression for the riskless interest rate $r_{f,t+1} = -\mathrm{E}_t m_{t+1} - \sigma_{mt}^2/2$, where $r_{f,t+1} \equiv \ln(1 + R_{f,t+1})$, $m_{t+1} \equiv \ln(M_{t+1})$, and $\sigma_{mt}^2 = \mathrm{Var}_t(m_{t+1})$. Applied to equation (4), it delivers an expression for the log risk premium, adjusted for Jensen's Inequality by adding one-half the own variance, $\mathrm{E}_t r_{i,t+1} - r_{f,t+1} + \sigma_i^2/2 = -\sigma_{imt}$, where $\sigma_{imt} \equiv \mathrm{Cov}_t(r_{i,t+1}, m_{t+1})$. In lognormal intertemporal equilibrium models, such as the representative-agent model with power utility, these equations are more convenient than equations (4) and (5).

A.5. Government Bond Returns

A largely separate literature has studied the behavior of the term structure of government bond yields. Until 1997, all U.S. Treasury bonds were nominal. Thus the great bulk of the literature studies the pricing of nominal bonds of different maturities. There are several important stylized facts.

First, the U.S. Treasury yield curve is upward-sloping on average. McCulloch and Kwon (1993), for example, report monthly zero-coupon bond yields that have been estimated from prices of coupon-bearing Treasury bonds. Using this data set and the sample period 1952:1 to 1991:2, Campbell et al. (1997, Chap. 10) report an average spread of 10-year zero-coupon log yields over one-month Treasury bill yields of 1.37 percent (137 basis points). This number can be taken as an estimate of the expected excess return on 10-year bonds if there is no expected upward or downward drift in nominal interest rates.[1]

Second, the U.S. Treasury yield curve is highly convex on average. That is, its average slope declines rapidly with maturity. The average yield spread over the one-month bill yield is 33 basis points at three months, 77 basis points at one year, and 96 basis points at two years. There is very little further change in average yields from two to 10 years.

Hansen and Jagannathan (1991) point out that the steep slope of the short-term Treasury yield curve implies high volatility of the SDF. The risk premia on longer-term Treasury bills over one- or three-month Treasury bills are small; but the volatility of excess returns in the Treasury bill market is also small, so Sharpe ratios are quite high. High Sharpe ratios of this sort, resulting from small excess returns divided by small standard deviations, are of course highly sensitive to transactions costs or liquidity services provided by Treasury bills. He and Modest (1995) and Luttmer (1996) show how to modify the basic Hansen–Jagannathan methodology to handle transactions costs and portfolio constraints, whereas Bansal and Coleman (1996) and Heaton and Lucas (1996) emphasize that liquidity services may depress Treasury bill returns relative to the returns on other assets.

A third stylized fact is that variations in U.S. Treasury yield spreads over time forecast future excess bond returns. This is true both at the short end of the term structure and at the long end, and it is true whether one measures a simple yield spread or the difference between a forward rate and a current spot rate (Shiller, Campbell, and Schoenholtz (1983), Fama and Bliss (1987), Campbell and Shiller (1991)). The predictability of excess bond returns contradicts the expectations hypothesis of the term structure, according to which expected excess bond returns are constant over time.

Just as in the literature on predictability of excess stock returns, it is important to keep in mind that the standard tests for unpredictability of returns may be subject to small-sample biases and peso problems. Small-sample biases arise because yield spreads are persistent and their innovations are correlated with bond returns, whereas peso problems arise if investors anticipate the possibility of a regime switch in interest rates that is not

---

[1]An alternative measure of the expected excess return is the realized average excess return over the sample period. In 1952 to 1991, this number is actually negative because nominal interest rates drifted upward over this period. Because there cannot be an upward drift in interest rates in the very long run, the realized average excess return over 1952 to 1991 is probably a downward-biased estimate of the term premium.

observed in the data sample. Bekaert, Hodrick, and Marshall (1998) consider both issues but conclude that there is indeed some genuine predictability of U.S. Treasury bond returns.

This evidence has implications for the relation between yield spreads and future movements in interest rates. If term premia are constant over time, then yield spreads are optimal predictors of future movements in interest rates. More generally, yield spreads contain predictions of both interest rates and term premia. In postwar U.S. data, short-term rates tend to increase when yield spreads are high, consistent with the expectations hypothesis, but long rates tend to fall, counter to that hypothesis (Campbell and Shiller (1991)). Looking across data sets drawn from different countries and time periods, yield spreads predict interest rate movements more successfully when the interest rate has greater seasonal or cyclical variation and less successfully when the monetary authority has smoothed the interest rate so that it follows an approximate random walk (Mankiw and Miron (1986), Hardouvelis (1994)).

B. Factor Structure of the SDF

The SDF can also be used to understand the enormous literature on multifactor models. Historically, this literature began with the insight of Sharpe (1964) and Lintner (1965), sharpened by Roll (1977), that if all investors are single-period mean-variance optimizers, then the market portfolio is mean-variance efficient, which implies a beta pricing relation between all assets and the market portfolio. Ross (1976) points out that this conclusion can also be reached using an asymptotic no-arbitrage argument and the assumption that the market portfolio is the only source of common, undiversifiable risk. More generally, if there are several common factors that generate undiversifiable risk, then a multifactor model holds.

Within the SDF framework, these conclusions can be reached directly from the assumption that the SDF is a linear combination of K common factors $f_{k,t+1}$, $k = 1 \ldots K$. For expositional simplicity I assume that the factors have conditional mean zero and are orthogonal to one another. If

$$M_{t+1} = a_t - \sum_{k=1}^{K} b_{kt} f_{k,t+1}, \tag{7}$$

then the negative of the covariance of any excess return with the SDF can be written as

$$-\mathrm{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1}) = \sum_{k=1}^{K} b_{kt}\sigma_{ikt} = \sum_{k=1}^{K} (b_{kt}\sigma_{kt}^2)\left(\frac{\sigma_{ikt}}{\sigma_{kt}^2}\right) = \sum_{k=1}^{K} \lambda_{kt}\beta_{ikt}.$$

$$\tag{8}$$

Here σ_{ikt} is the conditional covariance of asset return i with the kth factor, σ_{kt}^2 is the conditional variance of the kth factor, $\lambda_{kt} \equiv b_{kt}\sigma_{kt}^2$ is the "price of risk" of the kth factor, and $\beta_{ikt} \equiv \sigma_{ikt}/\sigma_{kt}^2$ is the "beta" or regression coefficient of asset return i on that factor. This equation, together with equation (5), implies that the risk premium on any asset can be written as a sum of the asset's betas with common factors times the risk prices of those factors.

This way of deriving a multifactor model is consistent with the earlier insights. In a single-period model with quadratic utility, for example, consumption equals wealth and the marginal utility of consumption is linear. In this case the SDF must be linear in future wealth, or equivalently linear in the market portfolio return. In a single-period model with K common shocks and completely diversifiable idiosyncratic risk, the SDF can depend only on the common shocks.

It is important to note the conditioning information in equation (8). Both the betas and the prices of risk are conditional on information at time t. Unfortunately, this conditional multifactor model does not generally imply an unconditional multifactor model of the same form. The relevant covariance for an unconditional model is the unconditional covariance $-\text{Cov}(M_{t+1}, R_{i,t+1} - R_{f,t+1}) = -\text{Cov}(a_t - \sum_{k=1}^{K} b_{kt}f_{k,t+1}, R_{i,t+1} - R_{f,t+1})$, and this involves covariances of the coefficients a_t and b_t with returns in addition to covariances of the factors $f_{k,t+1}$ with returns. One way to handle this problem is to model the coefficients themselves as linear functions of observable instruments: $a_t = a'z_t$ and $b_t = b'z_t$, where z_t is a vector of instruments including a constant. In this case one obtains an unconditional multifactor model in which the factors include the original $f_{k,t+1}$, the instruments z_t, and all cross-products of $f_{k,t+1}$ and z_t. Cochrane (1996) and Lettau and Ludvigson (1999b) implement this approach empirically, and Cochrane (1999, Chap. 7) provides a particularly clear explanation. Jagannathan and Wang (1996) develop a related approach including instruments as factors but excluding cross-product terms.

B.1. The Cross-Sectional Structure of Stock Returns

Early work on the Sharpe–Lintner Capital Asset Pricing Model (CAPM) tended to be broadly supportive. The classic studies of Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973), for example, found that high-beta stocks tended to have higher average returns than low-beta stocks and that the relation was roughly linear. Although the slope of the relation was too flat to be consistent with the Sharpe–Lintner version of the CAPM, this could be explained by borrowing constraints of the sort modeled by Black (1972).

During the 1980s and 1990s, researchers began to look at other characteristics of stocks besides their betas. Several deviations from the CAPM, or "anomalies," were discovered. First, Banz (1981) reported the *size effect* that small (low-market-value) stocks have higher average excess returns than can be explained by the CAPM. Small stocks do have higher betas and higher average returns than large stocks, but the relation between average return

and beta for size-sorted portfolios is steeper than the CAPM security market line. Fama and French (1992) drew further attention to the size effect by sorting stocks by both size and beta and showing that high-beta stocks have no higher returns than low-beta stocks of the same size. Second, several authors found a *value effect* that returns are predicted by ratios of market value to accounting measures such as earnings or the book value of equity (Basu (1983), Rosenberg, Reid, and Lanstein (1985), Fama and French (1992)). This is related to the finding of DeBondt and Thaler (1985) that stocks with low returns over the past three to five years outperform in the future. Third, Jegadeesh and Titman (1993) documented a *momentum effect* that stocks with high returns over the past three to 12 months tend to outperform in the future.

Empirically, these anomalies can be described parsimoniously using multi-factor models in which the factors are chosen atheoretically to fit the empirical evidence. Fama and French (1993) introduced a three-factor model in which the factors include the return on a broad stock index, the excess return on a portfolio of small stocks over a portfolio of large stocks, and the excess return on a portfolio of high book-to-market stocks over a portfolio of low book-to-market stocks. Carhart (1997) augmented the model to include a portfolio of stocks with high returns over the past few months. These models broadly capture the performance of stock portfolios grouped on these characteristics, with the partial exception of the smallest value stocks.

There is considerable debate about the interpretation of these results. The first and most conservative interpretation is that they are entirely spurious, the result of "data snooping" that has found accidental patterns in historical data (Lo and MacKinlay (1990), White (2000)). Some support for this view, in the case of the size effect, is provided by the underperformance of small stocks in the 15 years since the effect was first widely publicized.

A second view is that the anomalies result from the inability of a broad stock index to proxy for the market portfolio return. Roll (1977) takes the extreme position that the CAPM is actually untestable, because any negative results might be due to errors in the proxy used for the market. In response to this, Stambaugh (1982) has shown that tests of the CAPM are insensitive to the addition of other traded assets to the market proxy, and Shanken (1987) has shown that empirical results can only be reconciled with the CAPM if the correlation of the proxy with the true market is quite low.

Recent research in this area has concentrated on human capital, the present value of claims to future labor income. Because labor income is about two-thirds of U.S. GDP and capital income is only one-third of GDP, it is clearly important to model human capital as a component of wealth. Jagannathan and Wang (1996) argue that labor income growth is a good proxy for the return to human capital and find that the inclusion of this variable as a factor reduces evidence against the CAPM. In a similar spirit, Liew and Vassalou (2000) show that excess returns to value stocks help to forecast GDP growth, and Vassalou (1999) introduces GDP forecast revisions as an additional risk factor in a cross-sectional model.

A third view is that the anomalies provide genuine evidence against the CAPM but not against a broader rational model in which there are multiple risk factors. Fama and French (1993, 1996) have interpreted their three-factor model as evidence for a "distress premium"; small stocks with high book-to-market ratios are firms that have performed poorly and are vulnerable to financial distress (Chan and Chen (1991)), and they command a risk premium for this reason.

Fama and French do not explain why distress risk is priced, that is, why the SDF contains a distress factor. Given the high price of distress risk relative to market risk, this question cannot be ignored; in fact MacKinlay (1995) expresses skepticism that any rational model with omitted risk factors can generate sufficiently high prices for those factors to explain the cross-sectional pattern of stock returns.

One possibility is that the distress factor reflects the distinction between a conditional and unconditional asset pricing model. The CAPM may hold conditionally but fail unconditionally. If the risk premium on the market portfolio moves over time, and if the market betas of distressed stocks are particularly high when the market risk premium is high, then distressed stocks will have anomalously high average returns relative to an unconditional CAPM even if they obey a conditional CAPM exactly. Jagannathan and Wang (1996) try to capture this by using a yield spread between low- and high-quality bonds as an additional risk factor proxying for the market risk premium. Cochrane (1996) and Lettau and Ludvigson (1999b) introduce additional risk factors by interacting the market return with the dividend-to-price ratio and long-short yield spread, and a consumption-wealth-income ratio. These approaches reduce deviations from the model, and Lettau and Ludvigson are particularly successful in capturing the value effect. Campbell and Cochrane (2000) take a more theoretical approach, showing that a model with habit formation in utility, of the sort described in Section IV below, implies deviations from an unconditional CAPM of the magnitude found in the data even though the CAPM holds conditionally.

Alternatively, the CAPM may fail even as a conditional model, but the data may be described by an intertemporal CAPM of the sort proposed by Merton (1973). In this case additional risk factors may be needed to capture time variation in investment opportunities that are of concern to long-term investors. This possibility is discussed further in Section IV.

A fourth view is that the anomalies do not reflect any type of risk but are "mistakes" that disappear once market participants become aware of them. Keim (1983) pointed out that the small-firm effect was entirely attributable to excess returns on small firms in the month of January. A seasonal excess return of this sort is very hard to relate to risk, and if it is not purely the result of data snooping it should be expected to disappear once it becomes well-known to investors. Indeed the January effect does seem to have diminished in recent years.

The most radical view is that the anomalies reflect enduring psychological biases that lead investors to make irrational forecasts. Lakonishok, Shleifer, and Vishny (1994) argue that investors irrationally extrapolate past earn-

ings growth and thus overvalue companies that have performed well in the past. These companies have low book-to-market ratios and subsequently underperform once their earnings growth disappoints investors. Supporting evidence is provided by La Porta (1996), who shows that earnings forecasts of stock market analysts fit this pattern, and by La Porta et al. (1997), who show that the underperformance of stocks with low book-to-market ratios is concentrated on earnings announcement dates. This view has much in common with the previous one and differs only in predicting that anomalies will remain stable even when they have been widely publicized.

All these views have difficulties explaining the momentum effect. Almost any model in which discount rates vary can generate a value effect: stocks whose discount rates are high, whether for rational or irrational reasons, have low prices, high book-to-market ratios, and high subsequent returns. It is much harder to generate a momentum effect in this way, and Fama and French (1996) do not attempt to give a rational risk-based explanation for the momentum effect. Instead they argue that it may be the result of data snooping or survivorship bias (Kothari, Shanken, and Sloan (1995)). Psychological models also have difficulties in that momentum arises if investors underreact to news. Such underreaction is consistent with evidence for continued high returns after positive earnings announcements (Bernard (1992)), but it is hard to reconcile with the overreaction implied by the value effect. Several recent attempts to solve this puzzle are discussed in Section VI.

II. Prices, Returns, and Cash Flows

A. Solving the Present Value Relation

The empirical work described in Section I takes the stochastic properties of asset returns as given and merely asks how the first moments of returns are determined from their second moments. Although this approach can be informative, ultimately it is unsatisfactory because the second moments of asset returns are just as endogenous as the first moments. The field of asset pricing should be able to describe how the characteristics of payoffs determine asset prices, and thus the stochastic properties of returns.

When an asset lasts for only one period, its price can be determined straightforwardly from equation (1). The difficulty arises when an asset lasts for many periods and particularly when it makes payoffs at more than one date. There are several ways to tackle this problem.

A.1. Constant Discount Rates

First, if an asset has a constant expected return, then its price is a linear function of its expected future payoffs. From the definition of return, $1 + R_{t+1} = (P_{t+1} + D_{t+1})/P_t$, if the expected return is a constant R, then

$$P_t = \frac{E_t(P_{t+1} + D_{t+1})}{1 + R}. \tag{9}$$

This model is sometimes called the martingale model or random walk model of stock prices because, even though the stock price itself is not a martingale in equation (9), the discounted value of a portfolio with reinvested dividends is a martingale (Samuelson (1965)). The expectational difference equation (9) can be solved forward. If one assumes that the expected discounted future price has a limit of zero, $\lim_{K \to \infty} E_t P_{t+K}/(1 + R)^K = 0$, then one obtains

$$P_t = E_t \sum_{i=1}^{\infty} \frac{D_{t+i}}{(1 + R)^i}. \tag{10}$$

The right-hand side of equation (10) is sometimes called the "fundamental value" of an asset price, although it is important to keep in mind that this expression holds only under the very special condition of a constant discount rate.

Models of "rational bubbles" (Blanchard and Watson (1982), Froot and Obstfeld (1991)) challenge the assumption just made that the expected future discounted price has a limit of zero. Such models entertain the possibility that future prices are expected to grow forever at the rate of interest, in which case a bubble term B_t that satisfies $B_t = E_t B_{t+1}/(1 + R)$ can be added to the right-hand side of equation (10). The theoretical conditions that allow bubbles to exist are extremely restrictive, however. Negative bubbles are ruled out by limited liability that puts a floor of zero on the price of an asset; this implies that a bubble can never start once an asset is trading, because $B_t = 0$ implies $B_{t+1} = 0$ with probability one (Diba and Grossman (1988)). General equilibrium considerations also severely limit the circumstances in which bubbles can arise (Santos and Woodford (1997)).

An important special case arises when the expected rate of dividend growth is a constant, $E_t(D_{t+1}/D_t) = (1 + G)$. In this case equation (10) simplifies to the Gordon (1962) growth model, $P_t = E_t D_{t+1}/(R - G)$. This formula relates prices to prospective dividends, the discount rate, and the expected dividend growth rate; it is widely used by both popular writers (Glassman and Hassett (1999)) and academic writers (Heaton and Lucas (1999)) to interpret variations in prices relative to dividends, such as the spectacular runup in aggregate stock prices at the end of the 1990s. A difficulty with such applications is that they assume changes in R and G that are ruled out by assumption in the Gordon model. Thus the Gordon formula can only provide a rough guide to the price variations that occur in a truly dynamic model.

At the end of the 1970s most finance economists believed that equation (10) was a good approximate description of stock price determination for at least the aggregate market. LeRoy and Porter (1981) and Shiller (1981) challenged this orthodoxy by pointing out that aggregate stock prices seem to be far more volatile than plausible measures of expected future dividends. Their work assumed that both stock prices and dividends are stationary around a stochastic trend; Kleidon (1986) and Marsh and Merton (1986) responded that stock prices follow unit-root processes, but Campbell and Shiller (1988a, 1988b) and West (1988) found evidence for excess volatility even allowing for unit roots.

A.2. A Log-Linear Approximate Framework

Campbell and Shiller (1988a) extended the linear present-value model to allow for log-linear dividend processes and time-varying discount rates. They did this by approximating the definition of log return, $r_{t+1} = \log(P_{t+1} + D_{t+1}) - \log(P_t)$, around the mean log dividend-to-price ratio, $(\overline{d_t - p_t})$, using a first-order Taylor expansion. The resulting approximation is $r_{t+1} \approx k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t$, where ρ and k are parameters of linearization defined by $\rho \equiv 1/(1 + \exp(\overline{d_t - p_t}))$ and $k \equiv -\log(\rho) - (1 - \rho)\log(1/\rho - 1)$. When the dividend-to-price ratio is constant, then $\rho = P/(P + D)$, the ratio of the ex-dividend to the cum-dividend stock price. In the postwar quarterly U.S. data of Campbell (1999), the average price-to-dividend ratio has been 26.4 on an annual basis, implying that ρ should be about 0.964 in annual data.

Solving forward, imposing the "no-bubbles" terminal condition that $\lim_{j \to \infty} \rho^j(d_{t+j} - p_{t+j}) = 0$, taking expectations, and subtracting the current dividend, one gets

$$p_t - d_t = \frac{k}{1 - \rho} + \mathrm{E}_t \sum_{j=0}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}]. \tag{11}$$

This equation says that the log price-to-dividend ratio is high when dividends are expected to grow rapidly or when stock returns are expected to be low. The equation should be thought of as an accounting identity rather than a behavioral model; it has been obtained merely by approximating an identity, solving forward subject to a terminal condition, and taking expectations. Intuitively, if the stock price is high today, then from the definition of the return and the terminal condition that the dividend-to-price ratio is non-explosive, there must either be high dividends or low stock returns in the future. Investors must then expect some combination of high dividends and low stock returns if their expectations are to be consistent with the observed price.

Campbell (1991) extends this approach to obtain a decomposition of returns. Substituting equation (11) into the approximate return equation gives

$$r_{t+1} - \mathrm{E}_t r_{t+1} = (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}. \tag{12}$$

This equation says that unexpected stock returns must be associated with changes in expectations of future dividends or real returns. An increase in expected future dividends is associated with a capital gain today, whereas an increase in expected future returns is associated with a capital loss today. The reason is that with a given dividend stream, higher future returns can only be generated by future price appreciation from a lower current price.

Equation (12) can be used to understand the relation between the predictability of excess returns, described in Section I.A.4, and the excess volatility of prices and returns discussed in Section II.A.2. The equation implies that

the variance of unexpected returns is the variance of "dividend news" (the first term on the right-hand side), plus the variance of "expected return news" (the second term on the right-hand side), plus twice the covariance between the two types of news. Revisions in expected future returns move prices today and generate volatility in unexpected returns. The discounted sums on the right-hand side of equation (12) imply that this effect is greater, the more persistent the revisions in expected future returns; thus considerable extra volatility is created by predictability of returns from slow-moving variables such as dividend-to-price or market-to-book ratios.[2]

A.3. Alternatives to the Use of Dividends

Both the linear present value model with constant discount rates and the log-linear approximate model relate asset prices to expected future dividends. The use of dividends is not arbitrary but results from the accounting identity that links returns, prices, and dividends. Nonetheless the presence of dividends creates several difficulties for empirical work.

First, many companies pay cash to shareholders partly by repurchasing shares on the open market. This strategy is tax-advantaged and has become increasingly popular in recent years. Share repurchases do not invalidate any of the formulas given above, but they do make it essential to apply these formulas correctly. One can work on a per-share basis, measuring price per share and traditional dividends per share; in this case repurchases are not counted as dividends, but they do affect future dividends per share by reducing the number of shares. Alternatively, one can work on a total-value basis, measuring the market value of the company and the sum of its cash payments to shareholders, including both traditional dividends and repurchases. Liang and Sharpe (1999) emphasize that the exercise of executive stock options leads to share issues that must be set against repurchases. They estimate that the dividend yield on the S&P 500 index in the late 1990s rises by about one percent when net repurchases are added to traditional dividends.[3]

Second, many companies seem to be postponing the payment of dividends until much later in their life cycle. This means that current dividends are observed for fewer companies. Fama and French (1999) report that in 1993 to 1997, non-dividend-payers accounted for almost one-quarter of the value of the aggregate stock market, a tenfold increase in their value share since the mid-1960s. These firms must eventually pay cash to shareholders in some way, but it is clearly not fruitful to model their dividends as following a stable stochastic process.

In response to these difficulties, there is increasing interest in present-value models in which earnings, rather than dividends, are the driving variable. Ohlson (1990, 1995) assumes a constant discount rate and shows that

---

[2] One can understand a rational bubble as the limiting case where the price movement is so persistent (in fact, explosive) that it has no effect on the expected return at all.

[3] Cash paid by acquiring firms to shareholders of acquired firms is another form of indirect payout. The quantitative importance of this is hard to judge.

under clean-surplus accounting, the market value of a firm can be written as the book value of its equity, plus the discounted present value of future surplus earnings (earnings in excess of the discount rate times book value). Vuolteenaho (1999) combines this approach with the log-linear approximate framework, allowing for time-varying discount rates. He shows that cash flow news accounts for a greater fraction of return volatility for a typical stock than for the market as a whole, because cash flow news is largely idiosyncratic. Ang and Liu (1998) combine the Ohlson model with an explicit model of the SDF.

A.4. Exact Solution Methods

Although the log-linear framework of Section II.A.2 delivers useful insights, it does rely on an approximation of return. Two other approaches are available to get more accurate results. First, one can work with a discrete-state model in which dividend growth and returns, or the expectations of these variables, follow a Markov chain. Under this assumption one can solve for prices in each discrete state by solving a system of linear equations. This approach has been used by Mehra and Prescott (1985), Kandel and Stambaugh (1991), and others. The discrete-state model can be specified a priori or can be chosen using Gaussian quadrature methods to approximate an estimated continuous-state model.

A second exact approach treats an asset paying dividend at many future dates as the sum of many assets, each of which pays a dividend at a single future date. These single-payment or "zero-coupon" assets are often easier to model. This approach is most natural, and has been developed furthest, in the literature on fixed-income securities that I now describe.

B. Affine-Yield Models

B.1. Theoretical Structure

Consider a real zero-coupon bond that makes a single payment of one unit of consumption at time $t + n$. The bond price at time t, P_{nt}, must satisfy the general SDF relation equation (1), which in this special case becomes

$$P_{nt} = E_t[M_{t+1}P_{n-1,t+1}]. \tag{13}$$

This equation can be solved forward to obtain

$$P_{nt} = E_t[M_{t+1}M_{t+2}\ldots M_{t+n}]. \tag{14}$$

Just as the price of a single-period real bond is the expectation of the single-period SDF in equation (4), so the price of an n-period real bond is the expectation of the n-period SDF (the product of n successive single-period SDF's). This is natural because an n-period real bond is the riskless real

asset for an investor with a horizon of n periods (Modigliani and Sutch (1966)). Equation (14) makes it clear that a model of bond prices is equivalent to a time-series model of the SDF.

The yield on an n-period bond, Y_{nt}, is defined to be the discount rate that equates the bond's price to the present value of its future payment: $P_{nt} = 1/(1 + Y_{nt})^n$. Taking logs, $y_{nt} = -(1/n)p_{nt}$, so the log bond yield is just a linear transformation of the log bond price. Affine-yield models have the property that all log bond yields are linear ("affine") in a set of state variables describing the movements of the SDF. This property holds only under restrictive conditions on the SDF. Classic papers by Vasicek (1977) and Cox, Ingersoll, and Ross (1985) present simple examples; recent work has sought to characterize more generally the conditions on the SDF required for an affine-yield structure. Most of this work is set in continuous time and is surveyed by Sundaresan (2000). In the discrete-time framework used here, if we define a vector z_t containing the log SDF m_t and all state variables relevant for predicting its future values m_{t+i}, then an affine-yield model requires that both the conditional mean and the conditional variance-covariance matrix of z_{t+1} should be linear in z_t. The model also requires some restrictions on the conditional distribution of z_{t+1}; a normal distribution is sufficient but not necessary (Backus, Foresi, and Telmer (1998)). The simplest example is a univariate homoskedastic normal AR(1) process for m_{t+1}, which gives a discrete-time version of the Vasicek (1977) model of the term structure. A normal AR(1) process for m_{t+1} with a variance proportional to the level of m_{t+1} gives a discrete-time version of the Cox et al. (1985) model. In continuous time the affine-yield restrictions are similar. It is possible to allow for jumps in the state vector provided that the jump intensity is itself linear in the state variables.

Dai and Singleton (2000) present a careful analysis of the substantive restrictions imposed by the affine-yield structure. There are two basic issues. First, the affine-yield model limits the way in which volatility (driven either by a diffusion term or by jump intensity in continuous time) can vary with state variables. Variances can be constant or, more generally, linear in state variables but cannot vary nonlinearly. Second, the implied variance-covariance matrix of the state vector must be positive definite. This places few restrictions on the correlations of state variables when volatility is constant but is much harder to ensure when volatility is time-varying.

The affine-yield structure makes it possible to solve recursively for the unknown coefficients relating log bond yields to state variables. Starting from the terminal condition that the bond price at maturity is one (so the log bond price is zero), one can work out the term structure of interest rates to price bonds of arbitrary maturity. Coupon bonds can then be priced, if desired, as packages of zero-coupon bonds, one for each date at which a coupon payment is made.

The affine-yield structure can also be extended to study assets with uncertain future payoffs. There are three ways to do this. First, one can assume that the expected asset returns are unaffected by payoff uncertainty, which will be the case if the payoff uncertainty is uncorrelated with the SDF.

Gibbons and Ramaswamy (1993), for example, assume that inflation uncertainty is uncorrelated with the SDF, so inflation risk is unpriced, and study the real returns on nominal bonds. Second, one can change the numeraire to units in which future payoffs are known. Nominal Treasury bonds, for example, have uncertain real payoffs but certain nominal payoffs. One can define a nominal SDF, the real SDF times the ratio of the price level today to the price level tomorrow, and then use an affine model of the nominal SDF. The advantage of this approach is its simplicity; the disadvantage is that it is hard to relate the nominal SDF to the real risks that affect investors. Third, one can explicitly model the time-series process for the payoffs in relation to the SDF. If the state vector includes those variables that determine the payoffs, and if the previous assumptions on the state vector continue to hold, then the log asset price continues to be linear in the log state vector and the advantages of the affine framework are preserved. Campbell and Viceira (2000) apply this approach to nominal bonds; Bakshi and Chen (1997), Ang and Liu (1998), Bekaert and Grenadier (1999), and Berk, Green, and Naik apply it to stocks.

B.2. Empirical Applications to U.S. Treasury Bonds

Section I.A.5 reported several stylized facts about U.S. Treasury bond yields. Both the shape of the average yield curve and the time variation of term premia are inconsistent with single-factor affine-yield models of the nominal term structure. In single-factor models, the convexity of the average yield curve is related to the rate at which shocks to the short-term interest rate die out; but time-series evidence implies that interest-rate shocks die out much more slowly than would be implied by the rapidly declining slope of the average yield curve (Gibbons and Ramaswamy (1993)). As for term premia, they are counterfactually constant in a single-factor Vasicek (1977) model with constant volatility. A single-factor Cox et al. (1985) model with volatility linear in the level of the short-term interest rate can generate time variation in term premia, but in this model the yield spread is inversely related to volatility. If term premia are positive on average, they move positively with volatility so they move inversely with the yield spread, contrary to the empirical evidence (Backus et al. (1998)).

These problems have led to an explosion of research on more complex models of the nominal term structure. These include multifactor affine models that include both persistent state variables to fit the slowly decaying autocorrelations of interest rates and transitory state variables to fit the rapidly declining slope of the average yield curve; nonaffine models that allow for a nonlinear effect of the level of the interest rate on its volatility and drift; and regime-switching models that allow for occasional discrete changes in interest-rate behavior. Practitioners have gone in a somewhat different direction, introducing time-dependent parameters into simple affine models, but this does not build the understanding of the underlying economic structure that is the ultimate goal of academic research. Most of this work is set in continuous time and is reviewed by Sundaresan (2000).

Several issues remain unresolved by this research. First, real-interest-rate risk and inflation risk are conceptually distinct and may have very different risk prices. This distinction is blurred when term-structure models are fit directly to nominal interest rates. It can be addressed either by looking at inflation data jointly with nominal interest rates (Gibbons and Ramaswamy (1993), Campbell and Viceira (2000)) or by studying inflation-indexed bonds, which have been actively traded in the U.K. for 15 years (Brown and Schaefer (1994)) and have been issued in the United States since 1997. There is some evidence that both inflation risk and real-interest-rate risk are priced, so that the real term structure is upward sloping but flatter than the nominal term structure; however more work is needed on this subject.

Second, it is important to reconcile the characterization of the SDF provided by bond market data with the evidence from stock market data. Term-structure models of the SDF are ultimately unsatisfactory unless they can be related to the deeper general-equilibrium structure of the economy. Researchers often calibrate equilibrium models to fit stock market data alone, but this is a mistake because bonds carry equally useful information about the SDF. The short-term real interest rate is closely related to the conditional expected SDF and thus to the expected growth rate of marginal utility; in a representative-agent model with power utility of consumption, this links the real interest rate to expected consumption growth as discussed in Section IV.A. The risk premium on long-term bonds is also informative; it is small relative to the equity premium, and this remains true even if one divides by the standard deviation of return to calculate a Sharpe ratio. This fact is consistent with risk-based models of the equity premium in which the SDF is more highly correlated with stocks than with bonds, but it runs counter to models that explain the equity premium as the result of liquidity services from short-term Treasury debt that drive down the equilibrium short-term interest rate (Bansal and Coleman (1996)). Unless long-term Treasury bonds also provide liquidity services, the liquidity-based model should imply a large yield spread of long-term over short-term Treasury debt.

Third, the short-term nominal interest rate is an important driving variable in any nominal term-structure model. One cannot ignore the fact that the short rate is directly controlled by the monetary authority, which responds to macroeconomic conditions by changing the rate in discrete jumps. Balduzzi, Bertola, and Foresi (1997) and Piazzesi (1999) address this issue.

III. Consumption and Portfolio Choice for Long-Term Investors

The general theory outlined in Section I can also be used to understand the microeconomic theory of portfolio choice. In a one-period model, the connection is particularly clear. If an investor k lives off financial wealth and consumes all wealth tomorrow, then the investor's consumption tomorrow, $C_{k,t+1}$, is just the payoff $X^*_{k,t+1}$ on the optimal portfolio: $C_{k,t+1} = X^*_{k,t+1}$. Using the relation between the SDF and the growth rate of the marginal

utility of consumption, we have $M_{t+1} = \theta_t U_k'(X^*_{k,t+1})$, where θ_t known at time t captures time discounting and marginal utility of consumption at time t. Inverting this relation,

$$X^*_{k,t+1} = U_k'^{-1}(M_{t+1}/\theta_t). \tag{15}$$

Given an exogenous SDF that summarizes the available asset returns, this equation determines the payoffs on the optimal portfolio for an investor with utility function U_k. In the classic case where U_k is quadratic, U_k' is linear and the optimal portfolio can be found using the mean-variance analysis of Markowitz (1952). In a CAPM equilibrium where M_{t+1} is linear in the return on the market portfolio, an investor with linear U_k' holds a mix of a riskless asset and the market. If M_{t+1} is highly volatile, then the optimal portfolio payoff and consumption are also highly volatile unless the investor is highly risk averse (U_k' is rapidly declining).

The difficult task faced by modern portfolio theory is to extend these insights to a multiperiod setting in which investors seek to finance a stream of consumption over a long lifetime. Financial economists have understood for at least 30 years that the solution to a multiperiod portfolio choice problem can be very different from the solution to a static portfolio choice problem. Samuelson (1969) derived restrictive conditions under which these two solutions are the same (portfolio choice is "myopic"). For an investor with no labor income who maximizes the expected value of a time-separable von Neumann–Morgenstern utility function, sufficient conditions for myopic portfolio choice are either log utility (with constant relative risk aversion equal to one), or both power utility (constant relative risk aversion) and returns that are independently and identically distributed (IID) over time.

Merton (1969, 1971, 1973) shows that if investment opportunities are varying over time, then long-term investors generally care about shocks to investment opportunities—the productivity of wealth—and not just about wealth itself. They may seek to hedge their exposures to wealth productivity shocks, and this gives rise to intertemporal hedging demands for financial assets. Brennan, Schwartz, and Lagnado (1997) have coined the phrase "strategic asset allocation" to describe this farsighted response to time-varying investment opportunities.

Unfortunately the intertemporal consumption and portfolio choice problem is hard to solve in closed form. Cox and Huang (1989) made conceptual progress by showing that a version of equation (15) must hold for each future date in a multiperiod model, but this does not generally lead to a closed-form solution. Few solutions have been available outside those trivial cases where the Merton model reduces to the static model, severely limiting the empirical applicability of the Merton model and its influence on asset management practices. Very recently this situation has begun to change as a result of several related developments. First, computing power and numerical methods have advanced to the point at which realistic multi-

period portfolio choice problems can be solved numerically using discrete-state approximations. Second, financial theorists have discovered some new closed-form solutions to the Merton model. Third, approximate analytical solutions have been developed. These solutions are based on perturbations of known exact solutions. They offer analytical insights into investor behavior in models that fall outside the still limited class that can be solved exactly.

A. Specification of Utility

A first important issue is the specification of utility. Ross (1998) points out that the investment horizon can affect portfolio choice, even with IID returns, once we relax the assumption of power utility. However only a few utility specifications are consistent with the observation that interest rates and risk premia have remained stationary over more than a century of U.S. economic growth. Most models of utility imply that trends in per capita consumption and wealth would generate counterfactual trends in financial variables. For this reason most research works either with power utility or with generalizations that retain the scale-independence of power utility.[4]

One appealing generalization of power utility has been proposed by Epstein and Zin (1989, 1991) and Weil (1989), building on the work of Kreps and Porteus (1978). This model moves outside the von Neumann–Morgenstern framework of expected utility to separate the coefficient of relative risk aversion from the elasticity of intertemporal substitution in consumption. The Epstein–Zin objective function can be written as

$$U(C_t, \mathrm{E}_t(U_{t+1})) = \left[(1 - \delta) C_t^{1-\gamma/\theta} + \delta \left(\mathrm{E}_t(U_{t+1}^{1-\gamma}) \right)^{1/\theta} \right]^{\theta/1-\gamma}, \tag{16}$$

where C_t is consumption at time t, $\gamma > 0$ is the relative risk aversion coefficient, $\psi > 0$ is the elasticity of intertemporal substitution, $0 < \delta < 1$ is the time discount factor, $\theta \equiv (1 - \gamma)/(1 - 1/\psi)$, and $\mathrm{E}_t(\cdot)$ is the conditional expectation operator. For time-separable power utility, $\gamma = 1/\psi$ and hence $\theta = 1$. In this case the nonlinear recursive equation (16) becomes a linear equation that can be solved forward to deliver the familiar time-separable power utility function.

Epstein–Zin utility is harder to work with than time-separable power utility, but Epstein and Zin have used the assumption that the investor has no labor income and lives entirely off financial wealth to show that the intertemporal marginal rate of substitution takes the form $\delta(C_{t+1}/C_t)^{-\theta/\psi} (1 + R_{p,t+1})^{\theta-1}$, where $R_{p,t+1}$ is the investor's portfolio return. If consump-

---

[4] Models of income risk or noise trading, however, sometimes assume constant absolute risk aversion for tractability. To prevent trending risk premia, such models can assume a trend in the number of risk-averse investors (equivalently a trend in the risk-bearing capacity of the economy) to counteract the trend in the scale of the economy. Campbell and Kyle (1993) is an example.

tion and returns are jointly lognormal, this implies that expected consumption growth is linear in the expected portfolio return with slope coefficient ψ, and that each asset's risk premium must equal a weighted average of the asset's covariance with the optimal portfolio return (as in the static CAPM) and optimal consumption growth (as in the consumption CAPM).

Giovannini and Weil (1989) have shown that in this model, portfolio choice is myopic when $\gamma = 1$ for any value of ψ. Thus Samuelson's conditions for myopic portfolio choice can be relaxed slightly in this framework. When $\psi = 1$, the consumption-to-wealth ratio is constant for any value of γ, but portfolio choice is affected by intertemporal hedging demand unless $\gamma = 1$. Campbell (1993, 1996) and Campbell and Viceira (1999, 2000) have used the Epstein–Zin model with $\psi = 1$ as a benchmark case and have derived approximate solutions that are accurate if ψ is sufficiently close to one. Numerical work suggests that accuracy is acceptable for ψ less than about three.

An alternative rationale for this form of preferences is provided by Maenhout (1999). He builds on the work of Anderson, Hansen, and Sargent (1999) to show that Epstein–Zin preferences may result from power utility, together with a preference for "robustness" (choices that deliver desirable outcomes in the face of model uncertainty, evaluated using the worst plausible model). Equivalent "stochastic differential utility" preferences have been formulated in continuous time by Duffie and Epstein (1992) and have been used to study long-term portfolio choice by Schroder and Skiadas (1999).

It is possible to construct other models of utility that preserve the desirable scale-independence of power utility. One strategy is to write utility as a power function of some variable other than consumption: for example, consumption relative to a habit or subsistence level that captures the past history of consumption. This approach is discussed in Section IV.B. Another strategy is to model a reference point that grows with consumption or wealth and to define utility relative to this reference point in a scale-independent way. This approach is discussed in Section VI.B.1. For the rest of this section, however, I will follow the bulk of the microeconomic portfolio choice literature and consider investors who have power or Epstein–Zin utility.

B. Time-Varying Investment Opportunities

B.1. Variation in Real Interest Rates

One of the most obvious ways in which investment opportunities vary is that real interest rates change over time. Real affine-yield models surveyed in Section II.B capture this variation and derive implications for long-term bond yields. If there were no variation in real interest rates, inflation-indexed bonds would have constant prices and would be perfect substitutes for cash; thus any coherent model of the demand for long-term real bonds must be an intertemporal rather than a static model. Nominal interest rates are also influenced by inflation shocks, but here too an intertemporal model is appropriate to capture the effects of changing real interest rates.

Investment advisers often argue that long-term bonds are appropriate for conservative long-term investors. In fact, the recommended investment allocations of several financial columnists and mutual-fund companies summarized by Canner, Mankiw, and Weil (1997) systematically increase the ratio of bonds to stocks for more conservative investors. Bond returns are risky in the short term, however, so this advice contradicts the mutual-fund theorem of Tobin (1958), according to which all investors should hold the same portfolio of risky assets, the tangency portfolio of mean-variance analysis. Conservative investors should hold less of this portfolio and more of a short-term safe asset, but the relative proportions of different risky assets—including stocks and bonds—should be the same for conservative investors as for aggressive investors. Canner et al. call this contradiction the *asset allocation puzzle*.

The Tobin mutual-fund theorem takes a short-term perspective, defining a short-term safe asset as riskless and long-term bonds as risky. Such a perspective is inappropriate for a long-term investor who cares about the distribution of wealth at a fixed date in the distant future. Short-term assets are not safe for such an investor, because they must be rolled over at uncertain future interest rates. Instead, a long-term zero-coupon bond that matures on the terminal date is the safe asset for this investor (Modigliani and Sutch (1966)).

This insight has recently been extended to the more realistic case in which the investor cares about the stream of consumption that can be financed by wealth, rather than the distribution of wealth at any particular date. Brennan and Xia (1998) consider a power-utility investor with a finite horizon, able to buy inflation-indexed bonds at interest rates determined by the Vasicek (1977) model. They show that as risk aversion increases, the investor's portfolio converges to an inflation-indexed coupon bond maturing at the horizon. Wachter (1998) proves a more general version of this result.

Campbell and Viceira (2000) fit a two-factor homoskedastic model of the SDF to data on U.S. nominal interest rates, stock returns, and inflation. They consider an infinitely lived investor with Epstein–Zin utility, derive a portfolio solution that is exact when $\psi = 1$ and approximate otherwise, and show that the share of nominal bonds relative to stocks increases with risk aversion when bonds are inflation-indexed or when inflation risk is moderate as it has been in the United States in the last 15 years. In this way they resolve the asset allocation puzzle.

B.2. Variation in Risk Premia

It is often argued that stocks are relatively safer for long-term investors (Siegel 1999). If asset returns are IID, this cannot be true because the means and variances of all asset returns increase in proportion with the investment horizon. Thus any claim of this sort must be based on predictability of stock returns. In fact, the empirical evidence summarized in Section I.A.4 does imply that the variance of stock returns increases more slowly than proportionally with the investment horizon. Poterba and Summers (1988) present direct evidence for this effect based on the predominantly negative

univariate autocorrelations of stock returns, and Campbell (1991) reports indirect evidence based on the predictability of stock returns from the dividend-to-price ratio and the negative correlation between innovations to the stock return and the dividend-to-price ratio. Predictability of this sort, that reduces long-term risk relative to short-term risk, is often called mean reversion.

If stock returns are mean-reverting, this again implies that investment opportunities are time-varying. Once again an intertemporal model is needed to calculate optimal investment allocations. Kim and Omberg (1996) solve the portfolio choice problem of a long-lived investor, with power utility defined over wealth at a distant future date, who faces a constant riskless interest rate and a time-varying equity premium that follows a homoskedastic AR(1) process. They show that when the stock return is negatively correlated with innovations to the equity premium, then the investor increases the average allocation to stocks if risk aversion is greater than one. Wachter (1999) derives a similar result for an investor who has power utility defined over consumption, in the special case where the innovations to the stock return and the equity premium are perfectly correlated. Campbell and Viceira (1999) consider an infinitely lived investor with Epstein–Zin utility defined over consumption. Their solution is exact when $\psi = 1$ and approximate otherwise. They estimate the parameters of the model from postwar U.S. data and find that the estimated mean reversion dramatically increases the average optimal equity allocation of a conservative long-term investor.

In all these models the investor optimally times the market, altering the allocation to stocks as the equity premium changes. Thus the models do not support the buy-and-hold policy recommended by Siegel (1999) for long-term investors.[5] This also raises the issue of consistency with general equilibrium; with a constant supply of stocks not all investors can buy or sell stocks at the same time, so these models cannot be used to describe the behavior of a representative agent in general equilibrium. Models with time-varying risk aversion, of the sort described in Section IV.B, solve the complementary problem of finding preferences that make a representative agent content to buy and hold the market in the face of a time-varying equity premium.

Brennan, Schwartz, and Lagnado (1996), Campbell, Chan, and Viceira (1999), and Lynch (1999) combine time-varying real interest rates with time-varying risk premia. They estimate more complex empirical models in which both bonds and stocks can be used to hedge against changes in investment opportunities. The estimated predictability of stock returns tends to imply that hedging demand is largest for stocks. Brandt (1999) uses a nonparametric approach, avoiding the specification of a parametric model, and obtains generally similar results. Balduzzi and Lynch (1999) show how transactions costs can affect optimal long-term portfolios.

---

[5] A buy-and-hold policy may be a constrained optimum if investors are unable to leverage their equity holdings. In this case, however, the constrained problem must be solved from first principles because the constraint affects the optimal consumption policy in addition to the optimal portfolio.

All these models assume that investors know the parameters of the model for asset returns. This is clearly unrealistic given the substantial uncertainty among financial economists about these parameters. Bawa, Brown, and Klein (1979) provided an early treatment of parameter uncertainty in a short-horizon model with IID asset returns, and Kandel and Stambaugh (1996) have considered the effects on a short-horizon investor of uncertainty about return predictability. Williams (1977), Dothan and Feldman (1986), and Gennotte (1986) point out that in a long-horizon setting, optimal portfolios are affected by the fact that investors can learn about parameters over time. Brennan (1998) and Barberis (2000) explore this issue empirically. If an investor learns about the mean stock return by observing realized returns, then a positive stock return will lead the investor to revise upward his estimate of the mean stock return. This makes stock returns positively correlated with expected future stock returns, the opposite of the mean-reversion effect. Hedging demand is correspondingly reversed, and the stock demand of conservative long-term investors is reduced. Xia (1999) considers the possibility of learning about the degree of mean-reversion itself. Future research will undoubtedly push further in the direction of integrating the estimation and portfolio choice problems; Brandt (1999) shows one way to do this by directly estimating an optimal portfolio rule rather than a process for asset returns.

C. Labor Income and Portfolio Choice

Labor income can also have important effects on portfolio choice for long-term investors. An important paper on this topic is Bodie, Merton, and Samuelson (hereafter BMS) (1991). These authors show that exogenous, riskless labor income is equivalent to an implicit holding of riskless assets. Previous results apply to total asset holdings, so riskless labor income tilts explicit asset holdings toward risky assets. Exogenous labor income that is perfectly correlated with risky assets, on the other hand, is equivalent to an implicit holding of risky assets and tilts the financial portfolio toward safe assets. BMS also consider the possibility that investors can vary their labor supply endogenously. They find that this increases the willingness to take risks because investors can absorb financial losses both by cutting consumption and by adjusting labor supply.

BMS do not consider idiosyncratic risk in labor income that cannot be hedged using financial assets. Recent theoretical work shows that such background risk can have important effects on consumption and portfolio decisions; Gollier (2000) provides a textbook treatment. Here too it is extremely challenging to extend two-period results to a more realistic multiperiod setting. Viceira (1999) shows that lognormally distributed labor income risk, uncorrelated with financial asset risk, reduces the tilt toward risky assets but does not reverse it. This type of risk also has a large effect in stimulating wealth accumulation through precautionary saving as emphasized by Carroll (1997).

Bertaut and Haliassos (1997), Heaton and Lucas (1997), Cocco, Gomes, and Maenhout (1998), and Storesletten, Telmer, and Yaron (1998) have explored the effects of realistically calibrated labor income risk on portfolio choice over the

life cycle. Because the ratio of labor income to wealth rises early in adult life and then gradually declines, the willingness to take equity risk follows a similar pattern. Dammon, Spatt, and Zhang (1999) have explored the life-cycle portfolio effects of the tax code and particularly the adjustment of the capital gains tax basis at death. They find that the tax code deters elderly investors from selling stocks, and thus increases their average allocation to equities.

IV. Equilibrium Models with a Representative Agent

The theory of intertemporal optimization relates the SDF to the marginal utility process for an unconstrained investor. This in turn restricts the consumption of such an investor. The research summarized in Section III uses this fact to solve for the optimal consumption and portfolio choice of investors with exogenously specified utility functions who face exogenous asset return processes (equivalently, an exogenous SDF).

Lucas (1978) pointed out that this logic can be reversed to derive the equilibrium SDF implied by an exogenous consumption process and utility specification. He assumed that the economy can be described by a representative investor, with a standard utility function, who consumes aggregate consumption. The first-order conditions of this investor determine the SDF. Asset prices can be determined explicitly by modeling payoffs jointly with the SDF, using methods described in Section II. Lucas modeled the aggregate stock market as paying a dividend equal to aggregate consumption; this is equivalent to the traditional assumption in finance that the stock market is the "market portfolio" of all wealth.

A. Three Puzzles

With the standard assumption of power utility, this approach leads to three puzzles. First, the average return on the stock market is too high to be readily explained by the model. The SDF implied by power utility is $\delta(C_{t+1}/C_t)^{-\gamma}$, the time discount factor times aggregate consumption growth raised to the power $-\gamma$. If we assume joint lognormality of consumption growth and asset returns for expositional convenience, the negative covariance of the log return on asset i with the log SDF becomes $\gamma\sigma_{ic}$, the coefficient of risk aversion times the covariance of asset i with log consumption growth. This should equal the risk premium on asset i. But empirically, U.S. consumption growth is very smooth (Campbell (1999) reports a standard deviation of about one percent in quarterly postwar U.S. data). Thus the covariance of risky assets with consumption growth cannot be large, no matter how highly correlated with consumption growth these assets may be. To explain the equity premium, the coefficient of risk aversion must be a much higher number than has traditionally been considered plausible. This is the *equity premium puzzle* of Mehra and Prescott (1985).

Second, the volatility of stock returns is too high to be readily explained by the model. Stock returns are driven by innovations to consumption growth, which affect both expected future dividends and discount rates. Expected log

dividend growth equals expected log consumption growth $E_t \Delta c_{t+1}$, and in a homoskedastic lognormal model with constant risk premia, the expected log stock return is a constant plus $\gamma E_t \Delta c_{t+1}$, risk aversion times expected consumption growth. Using the approximate log-linear formula for stock returns, equation (12), the implied return on the aggregate stock market, $r_{e,t+1}$, is

$$r_{e,t+1} - E_t r_{e,t+1} = (\Delta c_{t+1} - E_t \Delta c_{t+1}) + (1 - \gamma)(E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j \Delta c_{t+1+j}.$$

(17)

If consumption growth is IID (this appears to be approximately true in postwar U.S. data), then expected future consumption growth is constant and the unexpected stock return should equal current unexpected consumption growth. More generally, variations in expected future consumption growth cause offsetting variations in expected future dividend growth and expected future stock returns—the offset is exact if risk aversion $\gamma = 1$—and this makes it hard to generate large variations in current stock returns. Campbell (1999) calls this the *stock market volatility puzzle*.

One response to the equity premium puzzle is to accept the possibility that risk aversion might be higher than was traditionally considered reasonable. Our intuition about risk aversion has been built through thought experiments, but Kandel and Stambaugh (1991) point out that the results of such experiments are very sensitive to details of their specification, including particularly the size of the gamble one considers.

Within the power utility framework, increasing risk aversion has the unfortunate side effect of lowering the elasticity of intertemporal substitution (EIS). Although there is direct evidence that the EIS is fairly small (Hall (1988), Campbell and Mankiw (1989)—but see Attanasio and Weber (1993) and Beaudry and van Wincoop (1996) for somewhat larger estimates), extremely small values for the EIS imply that investors have an overpowering preference for a flat consumption path. Given the historical upward drift in consumption, this implies an extremely strong desire to borrow from the future. Unless this is offset by a low or even negative rate of time preference, the result is a counterfactually high real interest rate. Weil (1989) calls this the *risk-free rate puzzle*.

Epstein–Zin utility can be helpful in resolving the risk-free rate puzzle. It, like the model of habit formation discussed in Section IV.B, allows the EIS to differ from the reciprocal of risk aversion. Thus one can postulate high risk aversion to resolve the equity premium puzzle without driving the EIS to an unreasonably low value.

There is no such straightforward escape from the stock market volatility puzzle. The lognormal model with power utility explains variation in stock prices relative to dividends as resulting from predictable variation in future consumption growth that moves expected future dividends and real interest rates. Unfortunately, real interest rates are too stable to explain large swings

in stock prices (Campbell (1991)), and there is very little evidence that stock price-to-dividend ratios predict future consumption growth, dividend growth, or real interest rates (Campbell (1999)). This suggests that the volatility of stock returns must be explained by changes in the equity premium itself.

One way to generate time variation in the equity premium is from time variation in volatility. Schwert (1989) presents evidence that the conditional volatility of stock returns moves countercyclically along with forecasts of stock returns, and Bollerslev, Engle, and Wooldridge (1988) derive implications of exogenous movements in return volatility for the equity premium. Of course, it is desirable to derive movements in return volatility from underlying fundamentals; accordingly Abel (1988), Kandel and Stambaugh (1991), Veronesi (1999), and Whitelaw (2000) model heteroskedasticity or time-varying uncertainty about the consumption (dividend) process. A difficulty with these efforts is that the evidence for heteroskedasticity in aggregate consumption is fairly weak. This suggests that it is worthwhile to build a model in which risk aversion changes over time so that the price of risk, rather than the quantity of risk, is time-varying.

B. Habit Formation and Time-Varying Risk Aversion

Sundaresan (1989) and Constantinides (1990) have argued for the importance of habit formation, a positive effect of today's consumption on tomorrow's marginal utility of consumption.

Several modeling issues arise at the outset. Writing the period utility function as $U(C_t, X_t)$, where X_t is the time-varying habit or subsistence level, the first issue is the functional form for $U(\cdot)$. Abel (1990) has proposed that $U(\cdot)$ should be a power function of the ratio C_t/X_t, whereas most other researchers have used a power function of the difference $C_t - X_t$. The second issue is the effect of an agent's own decisions on future levels of habit. In standard "internal habit" models such as those in Constantinides (1990) and Sundaresan (1989), habit depends on an agent's own consumption and the agent takes account of this when choosing how much to consume. In "external habit" models such as those in Abel (1990) and Campbell and Cochrane (1999), habit depends on aggregate consumption that is unaffected by any one agent's decisions. Abel calls this "catching up with the Joneses." The third issue is the speed with which habit reacts to individual or aggregate consumption. Abel (1990) and Ferson and Constantinides (1991) make habit depend on one lag of consumption, whereas Sundaresan (1989), Constantinides (1990), Heaton (1995), and Campbell and Cochrane (1999) assume that habit reacts only gradually to changes in consumption.

The choice between ratio models and difference models of habit is important because ratio models have constant risk aversion whereas difference models have time-varying risk aversion. In Abel's (1990) ratio model, external habit adds a term to the equation describing the riskless interest rate but does not change the equation that describes the excess return of risky assets over the riskless interest rate. The effect on the riskless interest rate has to do with intertemporal substitution. Holding consumption today and

expected consumption tomorrow constant, an increase in consumption yesterday increases the marginal utility of consumption today. This makes the representative agent want to borrow from the future, driving up the real interest rate.

This instability of the riskless real interest rate is a fundamental problem for habit-formation models. Time-nonseparable preferences make marginal utility volatile even when consumption is smooth, because consumers derive utility from consumption relative to its recent history rather than from the absolute level of consumption. But unless the consumption and habit processes take particular forms, time-nonseparability also creates large swings in expected marginal utility at successive dates, and this implies large movements in the real interest rate. I now present an alternative specification in which it is possible to solve this problem and in which risk aversion varies over time.

Campbell and Cochrane (1999) build a model with external habit formation in which a representative agent derives utility from the difference between consumption and a time-varying subsistence or habit level. They assume that log consumption follows a random walk with mean g and innovation $\epsilon_{c,t+1}$. This is a fairly good approximation for U.S. data. The utility function of the representative agent is a time-separable power utility function, with curvature γ, of the difference between consumption C_t and habit X_t. Utility is only defined when consumption exceeds habit.

It is convenient to capture the relation between consumption and habit by the surplus consumption ratio S_t, defined by $S_t \equiv (C_t - X_t)/C_t$. The surplus consumption ratio is the fraction of consumption that exceeds habit and is therefore available to generate utility. If habit X_t is held fixed as consumption C_t varies, the local coefficient of relative risk aversion is $-Cu_{CC}/u_C = \gamma/S_t$, where u_C and u_{CC} are the first and second derivatives of utility with respect to consumption. Risk aversion rises as the surplus consumption ratio S_t declines, that is, as consumption approaches the habit level. Note that γ, the curvature parameter in utility, is no longer the coefficient of relative risk aversion in this model.

To complete the description of preferences, one must specify how the habit X_t evolves over time in response to aggregate consumption. Campbell and Cochrane suggest an AR(1) model for the log surplus consumption ratio, $s_t \equiv \log(S_t)$:

$$s_{t+1} = (1 - \varphi)\bar{s} + \varphi s_t + \lambda(s_t)\epsilon_{c,t+1}. \tag{18}$$

The parameter φ governs the persistence of the log surplus consumption ratio, whereas the "sensitivity function" $\lambda(s_t)$ controls the sensitivity of s_{t+1} and thus of log habit x_{t+1} to innovations in consumption growth $\epsilon_{c,t+1}$. This modeling strategy ensures that the habit process implied by a process for s_{t+1} always lies below consumption.

To derive the asset pricing implications of this model, one first calculates the SDF as

$$M_{t+1} = \delta\left(\frac{S_{t+1}}{S_t}\right)^{-\gamma}\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{19}$$

The SDF is driven by proportional innovations in the surplus consumption ratio and also by proportional innovations in consumption. If the surplus consumption ratio is only a small fraction of consumption, then small shocks to consumption can be large shocks to the surplus consumption ratio; thus the SDF can be highly volatile even when consumption is smooth. This volatility is itself time-varying because it depends on the level of the surplus consumption ratio.

The logic of Hansen and Jagannathan (1991) implies that the largest possible Sharpe ratio is given by the conditional standard deviation of the log SDF. This is $\gamma\sigma(1 + \lambda(s_t))$, so a sensitivity function that varies inversely with s_t delivers a time-varying, countercyclical Sharpe ratio.

The same mechanism helps to stabilize the riskless real interest rate. When the surplus consumption ratio falls, investors have an intertemporal-substitution motive to borrow from the future, but this is offset by an increased precautionary savings motive created by the volatility of the SDF. Campbell and Cochrane parameterize the model so that these two effects exactly cancel. This makes the riskless real interest rate constant, a knife-edge case that helps to reveal the pure effects of time-varying risk aversion on asset prices. With a constant riskless rate, real bonds of all maturities are also riskless, and there are no real term premia. Thus the equity premium is also a premium of stocks over long-term bonds.

When this model is calibrated to fit the first two moments of consumption growth, the average riskless interest rate, and the Sharpe ratio on the stock market, it also roughly fits the volatility, predictability, and cyclicality of stock returns. The model does not resolve the equity premium puzzle, because it relies on high average risk aversion, but it does resolve the stock market volatility puzzle.

It is important to understand the mechanism by which this resolution takes place. Any scale-independent model implies that wealth and consumption are cointegrated, so over sufficiently long horizons their growth rates must have the same volatility. Because the volatility of wealth seems to be far higher than the volatility of consumption in the short run, the question is how these volatilities are reconciled in the long run. The two possibilities are that consumption growth is positively autocorrelated, so the variance of consumption growth increases more than proportionally with the horizon, or that stock returns are negatively autocorrelated, so the variance of wealth increases less than proportionally with the horizon. The Campbell–Cochrane model assumes random walk consumption and implies negative autocorrelation of stock returns. The Constantinides (1990) model of habit formation, by contrast, assumes IID asset returns and implies positive autocorrelation of consumption growth.

Campbell and Cochrane (2000) develop cross-sectional implications of the habit-formation model. Because the model is driven by a single shock, the stock market return is almost perfectly correlated with the innovation to consumption. This implies that either a conditional consumption CAPM or a conditional version of the traditional CAPM provides a good conditional description of returns. Unconditionally, however, the consumption CAPM works

poorly because the SDF is a time-varying function of consumption growth. The market return partly captures this time variation so the unconditional CAPM works better. A scaled version of the CAPM works better still. These results apply to artificial data generated by the habit-formation model, not to real data, but they do suggest the importance of using conditional models to evaluate cross-sectional anomalies as in the recent work of Cochrane (1996), Jagannathan and Wang (1996), and Lettau and Ludvigson (1999b).

V. Equilibrium Models with Heterogeneous Agents

The previous section assumed that assets can be priced as if there is a representative agent who consumes aggregate consumption. An alternative view is that aggregate consumption is not an adequate proxy for the consumption of investors.

A. Heterogeneous Constraints

One simple explanation for a discrepancy between these two measures of consumption is that there are two types of agents in the economy: constrained agents who are prevented from trading in asset markets and simply consume their labor income each period, and unconstrained agents. The consumption of the constrained agents is irrelevant to the determination of equilibrium asset prices, but it may be a large fraction of aggregate consumption. Campbell and Mankiw (1989) argue that predictable variation in consumption growth, correlated with predictable variation in income growth, suggests an important role for constrained agents, whereas Mankiw and Zeldes (1991) and Brav, Constantinides, and Geczy (1999) use U.S. panel data to show that the consumption of stockholders is more volatile and more highly correlated with the stock market than the consumption of nonstockholders. Heaton and Lucas (1999) and Vissing-Jørgensen (1997) build general equilibrium models in which a limited fraction of the population participates in the stock market. They argue that an increased participation rate, along with reduced costs of diversification, may help to explain recent increases in stock prices. Such effects are likely to be even more important in countries with low stock market capitalization and concentrated equity ownership.

This view suggests an approach to asset pricing in which one assumes that there is a representative investor who holds the market portfolio but does not necessarily consume aggregate consumption. Consumption becomes unobservable and must be substituted out of an intertemporal model. Merton's (1973) intertemporal CAPM is the classic example of this approach. More recently Campbell (1993, 1996) has developed an empirical discrete-time version of this model in which priced factors are innovations to variables that forecast future returns on the market portfolio. For reasons discussed in Section III.B.2, the estimated mean reversion of aggregate stock market returns increases the demand for stocks by the representative investor if the coefficient of relative risk aversion is greater than one. This

increase in equity demand means that the risk aversion needed to fit the equity premium is large, in line with the consumption-based literature of Section IV, rather than small, as suggested by Friend and Blume (1975) using a static model that ignores mean reversion. This work deepens the equity premium puzzle by showing that the puzzle arises even if one ignores aggregate consumption data.

One might hope that an intertemporal CAPM of this sort, with additional priced risks beyond shocks to the market, could explain the cross-sectional anomalies in stock returns. This line of research is promising but is hampered by the Roll (1977) critique that aggregate wealth itself is very hard to measure.

B. Heterogeneous Income

It is also possible that utility-maximizing stock market investors are heterogeneous in important ways. If investors are subject to large idiosyncratic risks in their labor income and can share these risks only indirectly by trading a few assets such as stocks and Treasury bills, their individual consumption paths may be much more volatile than aggregate consumption. Even if individual investors have the same power utility function, so that any individual's consumption growth rate raised to the power $-\gamma$ would be a valid SDF, the aggregate consumption growth rate raised to the power $-\gamma$ may not be a valid SDF.

This problem is an example of Jensen's Inequality. Because marginal utility is nonlinear, the average of investors' marginal utilities of consumption is not generally the same as the marginal utility of average consumption. The problem disappears when investors' individual consumption streams are perfectly correlated with one another as they will be in a complete markets setting. Grossman and Shiller (1982) point out that it also disappears in a continuous-time model when the processes for individual consumption streams and asset prices are diffusions.

Recently Constantinides and Duffie (1996) have provided a simple framework within which the effects of heterogeneity can be understood. Constantinides and Duffie postulate an economy in which individual investors k have different consumption levels C_{kt}. The cross-sectional distribution of individual consumption is lognormal, and the change from time t to time $t + 1$ in individual log consumption is cross-sectionally uncorrelated with the level of individual log consumption at time t. All investors have the same power utility function with time discount factor δ and coefficient of relative risk aversion γ.

In this economy each investor's own intertemporal marginal rate of substitution is a valid SDF. Hence the cross-sectional average of investors' intertemporal marginal rates of substitution, M_{t+1}^*, is a valid SDF. However the marginal utility of the cross-sectional average of investors' consumption, M_{t+1}^{RA}, is *not* a valid SDF when marginal utility is nonlinear. This false SDF would be used incorrectly by an economist who ignores the aggregation problem in the economy.

The difference between the logs of these two variables is

$$m_{t+1}^* - m_{t+1}^{RA} = \frac{\gamma(\gamma + 1)}{2} \text{Var}_{t+1}^* \Delta c_{k,t+1}, \tag{20}$$

where Var_{t+1}^* denotes a cross-sectional variance measured at time $t + 1$. The time series of this difference can have a nonzero mean, helping to explain the risk-free rate puzzle, and a nonzero variance, helping to explain the equity premium puzzle. If the cross-sectional variance of log consumption growth is negatively correlated with the level of aggregate consumption, so that idiosyncratic risk increases in economic downturns, then the true SDF m_{t+1}^* will be more strongly countercyclical than the representative-agent SDF constructed using the same preference parameters; this has the potential to explain the high price of risk without assuming that individual investors have high risk aversion. Mankiw (1986) makes a similar point in a two-period model.

An important unresolved question is whether the heterogeneity we can measure has the characteristics that are needed to help resolve the asset pricing puzzles. In the Constantinides–Duffie model the heterogeneity must be large to have important effects on the SDF; a cross-sectional standard deviation of log consumption growth of 20 percent, for example, is a cross-sectional variance of only 0.04, and it is variation in this number over time that is needed to explain the equity premium puzzle. Interestingly, the effect of heterogeneity is strongly increasing in risk aversion because $\text{Var}_{t+1}^* \Delta c_{k,t+1}$ is multiplied by $\gamma(\gamma + 1)/2$ in equation (20). This suggests that heterogeneity may supplement high risk aversion but cannot altogether replace it as an explanation for the equity premium puzzle.

Cogley (1998) looks at consumption data and finds that measured heterogeneity has only small effects on the SDF. Lettau (1997) reaches a similar conclusion by assuming that individuals consume their income and calculating the risk-aversion coefficients needed to put model-based SDFs inside the Hansen–Jagannathan volatility bounds. This procedure is conservative in that individuals trading in financial markets are normally able to achieve some smoothing of consumption relative to income. Nevertheless Lettau finds that high individual risk aversion is still needed to satisfy the Hansen–Jagannathan bounds.

These conclusions may not be surprising given the Grossman–Shiller (1982) result that the aggregation problem disappears in a continuous-time diffusion model. In such a model, the cross-sectional variance of consumption is locally deterministic, and hence the false SDF M_{t+1}^{RA} correctly prices risky assets. In a discrete-time model the cross-sectional variance of consumption can change randomly from one period to the next, but in practice these changes are likely to be small. This limits the effects of consumption heterogeneity on asset pricing.

It is also important to note that idiosyncratic shocks are assumed to be permanent in the Constantinides–Duffie model. Heaton and Lucas (1996) calibrate individual income processes to microdata from the Panel Study of

Income Dynamics (PSID). Because the PSID data show that idiosyncratic income variation is largely transitory, Heaton and Lucas find that investors can minimize its effects on their consumption by borrowing and lending. This prevents heterogeneity from having any large effects on aggregate asset prices.

To get around this problem, several recent papers have combined heterogeneity with constraints on borrowing. Heaton and Lucas (1996) and Krusell and Smith (1997) find that borrowing constraints or large costs of trading equities are needed to explain the equity premium. Constantinides, Donaldson, and Mehra (1998) focus on heterogeneity across generations. In a stylized three-period overlapping generations model young agents have the strongest desire to hold equities because they have the largest ratio of labor income to financial wealth. If these agents are prevented from borrowing to buy equities, the equilibrium equity premium is large.

C. Heterogeneous Preferences

Heterogeneity in preferences may also be important. Several authors have recently argued that trading between investors with different degrees of risk aversion or time preference, possibly in the presence of market frictions or portfolio insurance constraints, can lead to time variation in the market price of risk (Dumas (1989), Grossman and Zhou (1996), Wang (1996), Sandroni (1997), Chan and Kogan (1999)). Intuitively, risk-tolerant agents hold more risky assets so they control a greater share of wealth in good states than in bad states; aggregate risk aversion therefore falls in good states, producing effects similar to those of habit formation.

VI. Behavioral Finance

Behavioral finance has been one of the most active areas in asset pricing during the 1990s. Shiller (1984) and Summers (1986) are two key early references; the field has now matured to the point where textbook treatments are possible (Shiller (1999), Shleifer (2000)).

Behavioral finance models contain two key ingredients. First, they postulate nonstandard behavior, driven by irrationality or nonstandard preferences, on the part of at least some investors. Ideally the postulated behavior is supported by experimental or empirical evidence. Second, they assume that rational investors with standard preferences are limited in their desire or ability to offset the asset demands of the first group of investors. This means that irrational expectations or nonstandard preferences affect the prices of financial assets.

These models cannot be tested using data on aggregate consumption or the market portfolio, because rational utility-maximizing investors neither consume aggregate consumption (some is accounted for by nonstandard investors) nor hold the market portfolio (instead they shift in and out of the stock market). This makes it hard to test behavioral models without having detailed information on the investment strategies of different market participants.

A. Limits to Arbitrage

In the first phase of this research, the description of nonstandard investors was rudimentary. These investors were called "noise traders." Their demands for risky assets were assumed to follow simple exogenous processes. Papers of this sort include Shiller (1984), De Long et al. (1990), Cutler, Poterba, and Summers (1991), and Campbell and Kyle (1993). The main emphasis of this work is on the factors that limit the ability of utility-maximizing investors to absorb the demands of noise traders at constant prices. (This absorption of demand is often called "arbitrage" but should not be confused with the technical use of the term to mean a riskless profit opportunity.)

The basic factor is risk, and this can be enough by itself. In Campbell and Kyle (1993), for example, markets are perfect and investors are infinitely lived, but utility maximizers charge a price for bearing fundamental risk when they are asked to buy more risky assets from noise traders. The effects of risk aversion are reinforced if investors have short horizons. De Long et al. (1990) assume an overlapping generations structure in which utility-maximizing investors are forced to sell their asset holdings after one period. In this case utility maximizers are exposed to price fluctuations caused by noise trading and also to fundamental risk, and their willingness to offset noise is correspondingly reduced. The wealth or capital of utility maximizers is also relevant. A shock to noise demand that moves prices away from fundamental value reduces the wealth of utility maximizers and may reduce their trading capacity. This factor is likely to be particularly important if utility-maximizing investors are financed by uninformed outside lenders, who react to losses by cutting lending (Shleifer and Vishny (1997), Xiong (1999)). This point appears highly relevant for understanding the events of fall 1998, particularly the collapse of the hedge fund Long Term Capital Management.

Because risk is the key factor limiting the ability of utility maximizers to offset noise-trader demands, noise traders cannot create diversifiable random variation in asset prices. The effects of noise trading must either be systematic—that is, correlated with pervasive common factors in the economy—or must be limited to isolated instances. Lee, Shleifer, and Thaler (1991), for example, argue that noise traders determine the premia and discounts on closed-end funds and that these premia tend to move with the returns on small stocks. Small-stock returns are a systematic risk factor in the Fama–French (1993) model, whether for fundamental reasons or because they reflect systematic shifts in noise trader demands. Shleifer (1986) and Wurgler and Zhuravskaya (1999) argue that noise traders bid up the prices of stocks when they are included in the S&P 500 index. Because only one or a few stocks are included in the index at any one time, this effect cannot be readily diversified away. Noise traders may similarly affect the prices of initial public offerings if only one or a few IPOs are available at any one time.

B. Modeling Nonstandard Behavior

Recent behavioral research has placed greater emphasis on the modeling of nonstandard investor behavior and has sought to explain the aggregate

predictability and cross-sectional patterns in stock returns summarized in Section I. Some models alter standard assumptions about preferences, whereas others assume particular forms of irrational expectations.

B.1. Prospect Theory

Preference-based behavioral models often work with the prospect theory of Kahneman and Tversky (1979). According to this theory, people do not judge outcomes on an absolute scale but compare outcomes with an initial reference point. Their objective function has a kink at the reference point, so risk aversion is locally infinite at that point. The objective function is concave for "gains" (outcomes above the reference point) but is convex for "losses" (outcomes below the reference point). All these properties are based on experimental evidence, but such evidence can be hard to interpret because experimental rewards and penalties are necessarily small and may not reveal people's attitudes toward larger risks. For this reason many applications modify the model to drop certain features, particularly the convexity of the objective function in the realm of losses. A related model of "first-order risk aversion," with kinked utility at a reference point but without convexity in the realm of losses, is developed by Epstein and Zin (1990) (see also Gul (1991)).

Several important issues arise in using prospect theory, and experimental evidence gives little guidance. A first issue is the time horizon: at what intervals do investors evaluate outcomes using the Kahneman–Tversky objective function? A second issue is the determination and updating of the reference point. A third issue is the argument of the objective function. In a standard intertemporal model, investors care about consumption rather than wealth. Many papers in the behavioral literature argue instead that wealth is the argument of the objective function. Some go further and claim that investors have a separate "mental account" for each asset, evaluating outcomes on an asset-by-asset basis.

Benartzi and Thaler (1995) argue that wealth-based prospect theory can explain the equity premium puzzle if investors evaluate their wealth frequently and update their reference points to current wealth levels. In this case the kink at the reference point is always relevant and makes investors, in effect, highly risk-averse. Epstein and Zin (1990) find a similar effect in their model of first-order risk aversion.

Barberis, Huang, and Santos (hereafter BHS) (2000) combine a standard power utility function in consumption with wealth-based prospect theory. They show that the Benartzi–Thaler model, with high but constant risk aversion, fails to explain the stock market volatility puzzle. Appealing to experimental evidence of Thaler and Johnson (1990), they argue that aversion to losses varies with past outcomes; past success reduces effective risk aversion as investors feel they are "gambling with house money." This creates a time-varying price of risk that explains aggregate stock market volatility in a similar manner to Campbell and Cochrane (1999). The BHS model has a lower aversion to consumption risk than the Campbell–

Cochrane model, because it generates risk-averse behavior not only from standard aversion to consumption fluctuations but also from direct aversion to wealth fluctuations.

Gomes (1999) explores the effects of wealth-based prospect theory on the demand for risky assets in partial equilibrium. Gomes shows that a decline in asset prices with fixed expected payoffs can actually reduce asset demands as increased risk aversion, caused by reduced wealth, outweighs higher expected returns. At a certain point, however, there is a discontinuous jump in asset demand as investors become risk-loving in the domain of losses. Gomes also shows that the kink in the Kahneman–Tversky objective function makes it optimal for investors to hold no stock when the equity premium is small; this is quite different from the standard result that a risk-averse investor should always take some amount of a favorable gamble, and it might help to explain the existence of investors who do not participate in the stock market.

Shefrin and Statman (1985) reported empirical evidence that many investors are reluctant to realize losses. This is a particular puzzle because the tax code favors realizing losses and delaying the realization of gains. Odean (1998) argues that prospect theory can explain this phenomenon. However such an argument requires a separate mental account for each stock and also requires that investors treat unrealized losses differently from realized ones (i.e., realization triggers evaluation of the objective function). These radical assumptions are not implied by prospect theory itself.

B.2. Irrational Expectations

A number of papers have explored the consequences of relaxing the assumption that investors have rational expectations and understand the behavior of dividend and consumption growth. In the absence of arbitrage, there exist positive state prices that can rationalize the prices of traded financial assets. These state prices equal subjective state probabilities multiplied by ratios of marginal utilities in different states. Thus given any model of utility, there exist subjective probabilities that produce the necessary state prices and in this sense explain the observed prices of traded financial assets. The interesting question is whether these subjective probabilities are sufficiently close to objective probabilities, and sufficiently related to known psychological biases in behavior, to be plausible.

Many of the papers in this area work in partial equilibrium and assume that stocks are priced by discounting expected future dividends at a constant rate. This assumption makes it easy to derive any desired behavior of stock prices directly from assumptions on dividend expectations. Barsky and De Long (1993), for example, assume that investors believe dividends to be generated by a doubly integrated process, so that the dividend growth rate has a unit root. These expectations imply that rapid dividend growth increases stock prices more than proportionally, so that the price-to-dividend ratio rises when dividends are growing strongly. If dividend growth is in fact sta-

tionary, then the high price-to-dividend ratio is typically followed by dividend disappointments, low stock returns, and reversion to the long-run mean price-to-dividend ratio. Under this assumption of stationary dividend growth, Barsky and DeLong's model produces overreaction of stock prices to dividend news. At the cross-sectional level, it explains the value effect. At the aggregate level, the model can account for the volatility puzzle and the predictability of stock returns.

It is harder to build a model of irrational expectations that explains the cross-sectional momentum effect together with the value effect. Three recent attempts to do this include Barberis, Shleifer, and Vishny (hereafter BSV) (1998), Daniel, Hirshleifer, and Subrahmanyam (hereafter DHS) (1998), and Hong and Stein (1999). BSV assume that investors entertain two different models of dividends. In the first model, dividend growth is negatively autocorrelated so dividends are mean-reverting, whereas in the second model, dividend growth is positively autocorrelated so dividends display trends. Investors never change either model but use Bayesian methods to update their priors about the probability that each model is true. BSV assume that dividends in fact follow a random walk, so both models are false; the mean-reverting model makes investors underreact to news, whereas the trending model makes them overreact. In this setting a string of good or bad news can induce investors to switch from the mean-reverting model to the trending model, so the model can produce both underreaction and overreaction depending on recent events.

The BSV model assumes a representative irrational investor. DHS (1998) and Hong and Stein (1999), by contrast, assume that there are multiple investors who receive both private signals and public information. In the DHS model, all investors are symmetrical and overconfident; they place too great a weight on their own private signal and too little on public information. Furthermore, investors react to public information that is consistent with their private signal by increasing their overconfidence but react to inconsistent public information by a smaller reduction in overconfidence; DHS call this "biased self-attribution." These assumptions imply that private information triggers short-run overreaction, which initially increases as investors react in a biased manner to subsequent public information. Only when the weight of public information becomes overwhelming do investors abandon their overconfident misvaluation. Thus the model produces both short-run momentum and long-run reversal.

Hong and Stein (1999) assume that there are two types of irrational investors. "Newswatchers" receive private signals about fundamental value, which diffuse gradually through the newswatching population. These investors form price expectations based on their signals but are imperfectly rational in that they do not learn from market prices. "Momentum investors" have no private information and trade on the basis of the most recent change in price. These investors are imperfectly rational in that they do not trade optimally based on the entire history of price changes. The interaction between these two groups produces both a momentum effect, as

private information gradually affects prices and is reinforced by momentum trading, and a value effect, as momentum investors drive prices beyond fundamental value.

Another potentially important form of irrationality is a failure to understand the difference between real and nominal magnitudes. Modigliani and Cohn (1979) argued that investors suffer from inflation illusion, in effect discounting real cash flows at nominal interest rates. Ritter and Warr (1999) and Sharpe (1999) argue that inflation illusion may have led investors to bid up stock prices as inflation has declined since the early 1980s. An interesting issue raised by this literature is whether misvaluation is caused by a high *level* of inflation (in which case it is unlikely to be important today) or whether it is caused by *changes* in inflation from historical benchmark levels (in which case it may contribute to high current levels of stock prices).

A limitation of all these models is that they do not consider general equilibrium issues. If the models are supposed to apply to individual assets, then the misvaluations they produce are diversifiable and so can easily be arbitraged away by rational investors. If, on the other hand, the models apply to the market as a whole, then it is important to consider the implications for consumption. Using for simplicity the fiction that dividends equal consumption, investors' irrational expectations about dividend growth should be linked to their irrational expectations about consumption growth. Interest rates are not exogenous but, like stock prices, are determined by investors' expectations. Thus it is significantly harder to build a general equilibrium model with irrational expectations.

To see how irrationality can affect asset prices in general equilibrium, consider first a static model in which log consumption follows a random walk with drift. Investors understand that consumption is a random walk, but they underestimate its drift. Such irrational pessimism lowers the average risk-free rate, increases the equity premium, and has an ambiguous effect on the price-to-dividend ratio. Thus pessimism has the same effects on asset prices as a low rate of time preference and a high coefficient of risk aversion, and it can help to explain both the risk-free rate puzzle and the equity premium puzzle (Hansen, Sargent, and Tallarini (1997)).

To explain the volatility puzzle, a more complicated model of irrationality is needed. Suppose now that log consumption growth follows an AR(1) process but that investors overestimate the persistence of this process. In this model the equity premium falls when consumption growth has been rapid and rises when consumption growth has been weak. This model, which can be seen as a general equilibrium version of Barsky and De Long (1993), fits the apparent cyclical variation in the market price of risk. One difficulty with this story is that it has strong implications for bond market behavior. Consumption growth drives up the riskless interest rate and the real bond premium even while it drives down the equity premium. Cecchetti, Lam, and Mark (1997) handle this problem by allowing the degree of investors' irrationality itself to be stochastic and time-varying.

VII. Conclusion

This review of the past 20 years of asset pricing research suggests three main developments. First, we now have a rich set of stylized facts about interest rates, aggregate stock prices, and cross-sectional patterns in stock returns. Although some of these facts may turn out to be the result of data snooping, most of them are likely to survive as the subject of legitimate interest. Second, we have developed a variety of models in response to these facts. Even though these models fall within the general outlines established by an earlier generation of theorists, the details are new and important. Third, asset pricing is concerned with the sources of risk and the economic forces that determine the rewards for bearing risk. Summers (1985) infamously compared financial economists with "ketchup economists" obsessed with the relative prices of different-sized bottles of ketchup. He alleged that "financial economists, like ketchupal economists . . . are concerned with the interrelationships between the prices of different financial assets. They ignore what seems to many to be the more important question of what determines the overall level of asset prices" (p. 634). This accusation may never have been fair, but it certainly does not describe the field today.

A number of important issues remain as challenges for the new millennium. Some of these grow out of the literature I have described, and others I have not touched on at all.

First, we have only a poor understanding of how transactions costs can affect asset prices. Amihud and Mendelson (1986) have argued that transactions costs significantly reduce the prices of small stocks and may help to explain the small-firm effect. He and Modest (1995) and Luttmer (1996) have extended the Hansen–Jagannathan volatility bounds to allow for transactions costs. The bigger challenge is to explain how transactions costs—both in adjusting portfolios and in adjusting consumption, particularly of durable goods and housing—affect the prices of assets in general equilibrium. Recent work by Constantinides (1986), Grossman and Laroque (1990), Heaton and Lucas (1996), Vayanos (1998), and others makes a start on this task, but we still need simple, tractable models that incorporate these frictions.

Second, there is tantalizing evidence that extra risk factors suggested by dynamic asset pricing models can help to explain cross-sectional patterns in stock returns. Some of these extra factors proxy for omitted components of wealth, notably human capital; others capture time variation in the underlying model for the SDF. More work is needed to refine our understanding of these factor models. Here it will be helpful to use information from government and corporate bond markets as well as from stock markets. Ultimate success will require a much deeper understanding of the relation between the SDF and the equilibrium of the real economy.

Despite the promise of such research, in my opinion it is unrealistic to hope for a fully rational, risk-based explanation of all the empirical patterns that have been discovered in stock returns. A more reasonable view is that

rational models of risk and return describe a long-run equilibrium toward which financial markets gradually evolve. Some deviations from such models can be quickly arbitraged away by rational investors; others are much harder to arbitrage and may disappear only after a slow process of learning and institutional innovation. The value effect, for example, may result in part from investors' irrational extrapolation of poor earnings growth—put another way, their reluctance to hold badly managed companies in declining industries. This effect may disappear only as mutual funds become available that disguise the identities of companies in a value portfolio.

If this view is correct, a third challenge for empirical research is to devise ways to track the changing strength of asset pricing patterns over time. Some effects, such as the January effect in small-stock returns or the earnings announcement effect, may disappear quite quickly after they become widely known. Others may be much more stable. Relatively little academic work has explored such issues systematically.

Fourth, an important function of asset markets is to enable investors to share risks. Despite the extraordinary financial innovation that has occurred in the last 30 years, many risk-sharing arrangements are still quite primitive. It is particularly hard for investors to share risks that affect large subgroups of the population, falling between the purely idiosyncratic risks covered by insurance contracts and the aggregate risks that can be hedged using bonds and stock indexes. An important role for academic research is to evaluate potential financial innovations, including financial instruments that are indexed to inflation, house prices, or components of national or world income. Shiller (1994) describes an agenda for exploring many of these issues.

Fifth, there are some fascinating public policy questions that relate to the field of asset pricing. Two important examples are the debate over proposals to invest Social Security funds in risky assets and the question of how the Treasury should optimally structure the public debt. As risk management issues confront both households and governments, asset pricing will remain an active and relevant field of economics.

REFERENCES

Abel, Andrew B., 1988, Stock prices under time-varying dividend risk: An exact solution in an infinite-horizon general equilibrium model, *Journal of Monetary Economics* 22, 375–393.

Abel, Andrew B., 1990, Asset prices under habit formation and catching up with the Joneses, *American Economic Review Papers and Proceedings* 80, 38–42.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Anderson, E., Lars P. Hansen, and Thomas J. Sargent, 1999, Risk and robustness in equilibrium, Working paper, University of Chicago.

Ang, Andrew, and Jun Liu, 1998, A generalized earnings model of stock valuation, Working paper, Stanford University.

Attanasio, Orazio, and Gulielmo Weber, 1993, Consumption growth, the interest rate, and aggregation, *Review of Economic Studies* 60, 631–649.

Backus, David, Silverio Foresi, Abon Mozumdar, and Liuren Wu, 1998, Predictable changes in yields and forward rates, NBER Working paper 6379.

Backus, David, Silverio Foresi, and Chris Telmer, 1998, Discrete-time models of bond pricing, NBER Working paper 6736.

Baker, Malcolm, and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *Journal of Finance*, forthcoming.

Bakshi, Gurdip, and Zhiwu Chen, 1997, Stock valuation in dynamic economies, Working paper, University of Maryland.

Balduzzi, Pierluigi, Giuseppe Bertola, and Silverio Foresi, 1997, A model of target changes and the term structure of interest rates, *Journal of Monetary Economics* 39, 223–249.

Balduzzi, Pierluigi, and Anthony Lynch, 1999, Transaction costs and predictability: Some utility cost calculations, *Journal of Financial Economics* 52, 47–78.

Bansal, Ravi, and Wilbur J. Coleman II, 1996, A monetary explanation of the equity premium, term premium, and riskfree rate puzzles, *Journal of Political Economy* 104, 1135–1171.

Banz, Rolf W., 1981, The relation between return and market value of common stocks, *Journal of Financial Economics* 9, 3–18.

Barberis, Nicholas C., 2000, Investing for the long run when returns are predictable, *Journal of Finance* 55, 225–264.

Barberis, Nicholas C., Ming Huang, and Tano Santos, 2000, Prospect theory and asset prices, *Quarterly Journal of Economics*, forthcoming.

Barberis, Nicholas C., Andrei Shleifer, and Robert W. Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307–343.

Barsky, Robert, and J. Brad De Long, 1993, Why does the stock market fluctuate?, *Quarterly Journal of Economics* 107, 291–311.

Basu, Sanjoy, 1983, The relationship between earnings yield, market value, and return for NYSE common stocks: Further evidence, *Journal of Financial Economics* 12, 129–156.

Bawa, Vijay, Stephen Brown, and Roger Klein, 1979, *Estimation Risk and Optimal Portfolio Choice* (North-Holland, Amsterdam).

Beaudry, Paul, and Eric van Wincoop, 1996, The intertemporal elasticity of substitution: An exploration using a U.S. panel of state data, *Economica* 63, 495–512.

Bekaert, Geert, and Steven R. Grenadier, 1999, Stock and bond pricing in an affine economy, Working paper, Stanford University.

Bekaert, Geert, Robert J. Hodrick, and David A. Marshall, 1998, Peso problem explanations for term structure anomalies, NBER Working Paper 7346.

Benartzi, Shlomo, and Richard H. Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics* 110, 73–92.

Berk, Jonathan B., Richard C. Green, and Vasant Naik, 1999, Optimal investment, growth options, and security returns, *Journal of Finance* 54, 1553–1607.

Bernard, Victor, 1992, Stock price reactions to earnings announcements, in Richard Thaler, ed.: *Advances in Behavioral Finance* (Russell Sage Foundation, New York).

Bertaut, Carol C., and Michael Haliassos, 1997, Precautionary portfolio behavior from a life-cycle perspective, *Journal of Economic Dynamics and Control* 21, 1511–1542.

Black, Fischer, 1972, Capital market equilibrium with restricted borrowing, *Journal of Business* 45, 444–454.

Black, Fischer, Michael Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in Michael Jensen, ed.: *Studies in the Theory of Capital Markets* (Praeger, New York).

Blanchard, Olivier J., and Mark Watson, 1982, Bubbles, rational expectations, and financial markets, in Paul Wachtel, ed.: *Crises in the Economic and Financial Structure: Bubbles, Bursts, and Shocks* (Lexington, Lexington, Mass.).

Bodie, Zvi, Robert C. Merton, and William F. Samuelson, 1991, Labor supply flexibility and portfolio choice in a life cycle model, *Journal of Economic Dynamics and Control* 16, 427–449.

Bollerslev, Timothy, Robert F. Engle, and Jeffrey Wooldridge, 1988, A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.

Brandt, Michael, 1999, Estimating portfolio and consumption choice: A conditional Euler equations approach, *Journal of Finance* 54, 1609–1645.

Brav, Alon, George Constantinides, and Christopher C. Geczy, 1999, Asset pricing with hetero-
geneous consumers and limited participation: Empirical evidence, Working paper, Duke
University.

Brennan, Michael J., 1998, The role of learning in dynamic portfolio decisions, *European Fi-
nance Review* 1, 295–306.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1996, The use of Treasury bill
futures in strategic asset allocation programs, Working paper, Anderson Graduate School of
Management, UCLA.

Brennan, Michael J., Eduardo S. Schwartz, and Ronald Lagnado, 1997, Strategic asset alloca-
tion, *Journal of Economic Dynamics and Control* 21, 1377–1403.

Brennan, Michael J., and Yihong Xia, 1998, Resolution of a financial puzzle, Working paper,
Anderson Graduate School of Management, UCLA.

Brown, Roger H., and Stephen M. Schaefer, 1994, The term structure of real interest rates and
the Cox, Ingersoll, and Ross model, *Journal of Financial Economics* 35, 3–42.

Brown, Stephen, William Goetzmann, and Stephen A. Ross, 1995, Survival, *Journal of Finance*
50, 853–873.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics*
18, 373–399.

Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101,
157–179.

Campbell, John Y., 1993, Intertemporal asset pricing without consumption data, *American Eco-
nomic Review* 83, 487–512.

Campbell, John Y., 1996, Understanding risk and return, *Journal of Political Economy* 104,
298–345.

Campbell, John Y., 1999, Asset prices, consumption, and the business cycle, in John Taylor and
Michael Woodford, eds.: *Handbook of Macroeconomics*, Vol. 1 (North-Holland, Amsterdam).

Campbell, John Y., Yeung Lewis Chan, and Luis M. Viceira, 1999, A multivariate model of
strategic asset allocation, Working paper, Harvard University.

Campbell, John Y., and John H. Cochrane, 1999, By force of habit: A consumption-based expla-
nation of aggregate stock market behavior, *Journal of Political Economy* 107, 205–251.

Campbell, John Y., and John H. Cochrane, 2000, Explaining the poor performance of consumption-
based asset pricing models, *Journal of Finance*, forthcoming.

Campbell, John Y., and Albert S. Kyle, 1993, Smart money, noise trading, and stock price
behavior, *Review of Economic Studies* 60, 1–34.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial
Markets* (Princeton University Press, Princeton, N.J.).

Campbell, John Y., and N. Gregory Mankiw, 1989, Consumption, income, and interest rates:
Reinterpreting the time-series evidence, in O. J. Blanchard and S. Fischer, eds.: *National
Bureau of Economic Research Macroeconomics Annual* 4 (MIT Press, Cambridge, Mass.).

Campbell, John Y., and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of
future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y., and Robert J. Shiller, 1988b, Stock prices, earnings, and expected dividends,
Journal of Finance 43, 661–676.

Campbell, John Y., and Robert J. Shiller, 1991, Yield spreads and interest rates: A bird's eye
view, *Review of Economic Studies* 58, 495–514.

Campbell, John Y. and Luis M. Viceira, 1999, Consumption and portfolio decisions when ex-
pected returns are time varying, *Quarterly Journal of Economics* 114, 433–495.

Campbell, John Y., and Luis M. Viceira, 2000, Who should buy long-term bonds? *American
Economic Review*, forthcoming.

Canner, Niko, N. Gregory Mankiw, and David N. Weil, 1997, An asset allocation puzzle, *Amer-
ican Economic Review* 87, 181–191.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52,
57–82.

Carroll, Christopher D., 1997, Buffer-stock saving and the life cycle/permanent income hypoth-
esis, *Quarterly Journal of Economics* 112, 1–55.

Cecchetti, Stephen G., Pok-Sang Lam, and Nelson C. Mark, 1994, Testing volatility restrictions on intertemporal marginal rates of substitution implied by Euler equations and asset returns, *Journal of Finance* 49, 123–152.

Cecchetti, Stephen G., Pok-Sang Lam, and Nelson C. Mark, 1997, Asset pricing with distorted beliefs: Are equity returns too good to be true?, Working paper, Ohio State University.

Chan, K. C., and Nai-fu Chen, 1991, Structural and return characteristics of small and large firms, *Journal of Finance* 46, 1467–1484.

Chan, Yeung Lewis, and Leonid Kogan, 1999, Heterogeneous preferences and the dynamics of asset prices, Working paper, Harvard University.

Cocco, Joao, Francisco Gomes, and Pascal Maenhout, 1998, Consumption and portfolio choice over the life cycle, Working paper, Harvard University.

Cochrane, John H., 1996, A cross-sectional test of an investment-based asset pricing model, *Journal of Political Economy* 104, 572–621.

Cochrane, John H., 1999, *Asset Pricing*, Unpublished manuscript, University of Chicago.

Cochrane, John H., and Lars Peter Hansen, 1992, Asset pricing lessons for macroeconomics, in O. J. Blanchard and S. Fischer, eds.: *NBER Macroeconomics Annual 1992* (MIT Press, Cambridge, Mass.).

Cogley, Timothy, 1998, Idiosyncratic risk and the equity premium: Evidence from the Consumer Expenditure Survey, Working paper, Federal Reserve Bank of San Francisco.

Constantinides, George, 1986, Capital market equilibrium with transactions costs, *Journal of Political Economy* 94, 842–862.

Constantinides, George, 1990, Habit formation: A resolution of the equity premium puzzle, *Journal of Political Economy* 98, 519–543.

Constantinides, George, John Donaldson, and Rajnish Mehra, 1998, Junior can't borrow: A new perspective on the equity premium puzzle, NBER Working paper 6617.

Constantinides, George, and Darrell Duffie, 1996, Asset pricing with heterogeneous consumers, *Journal of Political Economy* 104, 219–240.

Cox, John C., and Chi-fu Huang, 1989, Optimal consumption and portfolio policies when asset prices follow a diffusion process, *Journal of Economic Theory* 39, 33–83.

Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Cox, John C., and Stephen A. Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145–166.

Cutler, David, James Poterba, and Lawrence Summers, 1991, Speculative dynamics, *Review of Economic Studies* 58, 529–546.

Dai, Qiang, and Kenneth J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance*, forthcoming.

Dammon, Robert M., Chester S. Spatt, and Harold H. Zhang, 1999, Optimal consumption and investment with capital gains taxes, Working paper, Carnegie Mellon University.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839–1885.

DeBondt, Werner F. M., and Richard H. Thaler, 1985, Does the stock market overreact?, *Journal of Finance* 40, 793–805.

De Long, J. Bradford, Andrei Shleifer, Lawrence Summers, and Michael Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703–738.

Diba, Behzad, and Herschel I. Grossman, 1988, The theory of rational bubbles in stock prices, *Economic Journal* 98, 746–757.

Dothan, M., and David Feldman, 1986, Equilibrium interest rates and multiperiod bonds in a partially observable economy, *Journal of Finance* 41, 369–382.

Duffie, Darrell, 1992, *Dynamic Asset Pricing Theory* (Princeton University Press, Princeton, N.J.).

Duffie, Darrell, and Lawrence Epstein, 1992, Stochastic differential utility and asset pricing, *Econometrica* 60, 353–394.

Dumas, Bernard, 1989, Two-person dynamic equilibrium in the capital market, *Review of Financial Studies* 2, 157–188.

Epstein, Lawrence, and Stanley Zin, 1989, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* 57, 937–969.

Epstein, Lawrence, and Stanley Zin, 1990, First-order risk aversion and the equity premium puzzle, *Journal of Monetary Economics* 26, 387–407.

Epstein, Lawrence, and Stanley Zin, 1991, Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical investigation, *Journal of Political Economy* 99, 263–286.

Fama, Eugene F., 1975, Short-term interest rates as predictors of inflation, *American Economic Review* 65, 269–282.

Fama, Eugene F., and Robert Bliss, 1987, The information in long-maturity forward rates, *American Economic Review* 77, 680–692.

Fama, Eugene F., and Kenneth R. French, 1988a, Permanent and temporary components of stock prices, *Journal of Political Economy* 96, 246–273.

Fama, Eugene F., and Kenneth R. French, 1988b, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–27.

Fama, Eugene F., and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Fama, Eugene F., and Kenneth R. French, 1999, Disappearing dividends: Changing firm characteristics or increased reluctance to pay?, Working paper, University of Chicago and MIT.

Fama, Eugene F., and J. MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 71, 607–636.

Ferson, Wayne E., and George Constantinides, 1991, Habit persistence and durability in aggregate consumption: Empirical tests, *Journal of Financial Economics* 29, 199–240.

Friend, Irwin, and Marshall E. Blume, 1975, The demand for risky assets, *American Economic Review* 65, 900–922.

Froot, Kenneth, and Maurice Obstfeld, 1991, Intrinsic bubbles: The case of stock prices, *American Economic Review* 81, 1189–1217.

Garcia, René, and Pierre Perron, 1996, An analysis of the real interest rate under regime shifts, *Review of Economics and Statistics* 78, 111–125.

Gennotte, Gerard, 1986, Optimal portfolio choice under incomplete information, *Journal of Finance* 41, 733–746.

Gibbons, Michael R., and Krishna Ramaswamy, 1993, A test of the Cox, Ingersoll, and Ross model of the term structure, *Review of Financial Studies* 6, 619–658.

Giovannini, Alberto, and Philippe Weil, 1989, Risk aversion and intertemporal substitution in the capital asset pricing model, NBER Working paper 2824.

Glassman, James K., and Kevin A. Hassett, 1999, *Dow 36,000: The New Strategy for Profiting from the Coming Rise in the Stock Market* (Times Books).

Glosten, Lawrence, Ravi Jagannathan, and David Runkle, 1993, On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance* 48, 1779–1801.

Gollier, Christian, 2000, *The Economics of Risk and Time* (MIT Press, Cambridge, Mass.), forthcoming.

Gomes, Francisco, 1999, Loss aversion and the demand for risky assets, Working paper, Harvard University.

Gordon, Myron, 1962, *The Investment, Financing, and Valuation of the Corporation* (Irwin, Homewood, Ill.).

Gray, Stephen F., 1996, Modelling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* 42, 27–62.

Grossman, Sanford J., and Guy Laroque, 1990, Asset pricing and optimal portfolio choice in the presence of illiquid durable consumption goods, *Econometrica* 58, 25–51.

Grossman, Sanford J., and Robert J. Shiller, 1981, The determinants of the variability of stock market prices, *American Economic Review* 71, 222–227.

Grossman, Sanford J., and Robert J. Shiller, 1982, Consumption correlatedness and risk measurement in economies with nontraded assets and heterogeneous information, *Journal of Financial Economics* 10, 195–210.

Grossman, Sanford J., and Z. Zhou, 1996, Equilibrium analysis of portfolio insurance, *Journal of Finance* 51, 1379–1403.

Gul, Faruk, 1991, A theory of disappointment aversion, *Econometrica* 59, 667–686.

Hall, Robert E., 1988, Intertemporal substitution in consumption, *Journal of Political Economy* 96, 221–273.

Hansen, Lars P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars P., John Heaton, and Erzo Luttmer, 1995, Econometric evaluation of asset pricing models, *Review of Financial Studies* 8, 237–274.

Hansen, Lars P., and Ravi Jagannathan, 1991, Restrictions on intertemporal marginal rates of substitution implied by asset returns, *Journal of Political Economy* 99, 225–262.

Hansen, Lars P., and Ravi Jagannathan, 1997, Assessing specification errors in stochastic discount factor models, *Journal of Finance* 52, 557–590.

Hansen, Lars P., and Scott Richard, 1987, The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models, *Econometrica* 55, 587–613.

Hansen, Lars P., Thomas J. Sargent, and Thomas Tallarini, 1997, Robust permanent income and pricing, Working paper, University of Chicago.

Hardouvelis, Gikas A., 1994, The term structure spread and future changes in long and short rates in the G7 countries: Is there a puzzle? *Journal of Monetary Economics* 33, 255–283.

Harrison, John M., and David Kreps, 1979, Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* 20, 381–408.

Harvey, Campbell R., 1989, Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.

Harvey, Campbell R., 1991, The world price of covariance risk, *Journal of Finance* 46, 111–157.

He, Hua, and David Modest, 1995, Market frictions and consumption-based asset pricing, *Journal of Political Economy* 103, 94–117.

Heaton, John, 1995, An empirical investigation of asset pricing with temporally dependent preference specifications, *Econometrica* 63, 681–717.

Heaton, John, and Deborah Lucas, 1996, Evaluating the effects of incomplete markets on risk sharing and asset pricing, *Journal of Political Economy* 104, 668–712.

Heaton, John, and Deborah Lucas, 1997, Market frictions, saving behavior and portfolio choice, *Macroeconomic Dynamics* 1, 76–101.

Heaton, John, and Deborah Lucas, 1999, Stock prices and fundamentals, *NBER Macroeconomics Annual* (MIT Press, Cambridge, Mass) 213–242.

Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *Review of Financial Studies* 5, 357–386.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Ingersoll, Jonathan E., Jr., 1987, *Theory of Financial Decision Making* (Rowman and Littlefield, Totowa, N.J.).

Jagannathan, Ravi, and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, *Journal of Finance* 51, 3–53.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jorion, Philippe, and William N. Goetzmann, 1999, Global stock markets in the twentieth century, *Journal of Finance* 54, 953–980.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263–291.

Kandel, Shmuel, and Robert F. Stambaugh, 1991, Asset returns and intertemporal preferences, *Journal of Monetary Economics* 27, 37–91.

Kandel, Shmuel, and Robert F. Stambaugh, 1996, On the predictability of stock returns: An asset allocation perspective, *Journal of Finance* 51, 385–424.

Keim, Donald, 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13–32.

Keim, Donald, and Robert Stambaugh, 1986, Predicting returns in stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kim, Tong Suk, and Edward Omberg, 1996, Dynamic nonmyopic portfolio behavior, *Review of Financial Studies* 9, 141–161.

Kleidon, Allan, 1986, Variance bounds tests and stock price valuation models, *Journal of Political Economy* 94, 953–1001.

Kothari, S. P., Jay Shanken, and Richard G. Sloan, 1995, Another look at the cross-section of expected stock returns, *Journal of Finance* 50, 185–224.

Kreps, David M., and E. L. Porteus, 1978, Temporal resolution of uncertainty and dynamic choice theory, *Econometrica* 46, 185–200.

Krusell, Per, and A. A. Smith, Jr., 1997, Income and wealth heterogeneity, portfolio choice, and equilibrium asset returns, Working paper, University of Rochester and Carnegie Mellon University.

Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny, 1994, Contrarian investment, extrapolation, and risk, *Journal of Finance* 49, 1541–1578.

Lamont, Owen, 1998, Earnings and expected returns, *Journal of Finance* 53, 1563–1587.

La Porta, Rafael, 1996, Expectations and the cross-section of returns, *Journal of Finance* 51, 1715–1742.

La Porta, Rafael, Josef Lakonishok, Andrei Shleifer, and Robert W. Vishny, 1997, Good news for value stocks: Further evidence on market efficiency, *Journal of Finance* 52, 859–874.

Lee, Charles M. C., Andrei Shleifer, and Richard Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75–110.

LeRoy, Stephen, and Richard Porter, 1981, The present value relation: Tests based on variance bounds, *Econometrica* 49, 555–577.

Lettau, Martin, 1997, Idiosyncratic risk and volatility bounds, Working paper, CentER, Tilburg University.

Lettau, Martin, and Sydney Ludvigson, 1999a, Consumption, aggregate wealth, and expected stock returns, Working paper, Federal Reserve Bank of New York.

Lettau, Martin, and Sydney Ludvigson, 1999b, Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying, Working paper, Federal Reserve Bank of New York.

Lewellen, Jonathan, 1999, The time-series relations among expected return, risk, and book-to-market, *Journal of Financial Economics* 54, 5–43.

Liang, J. Nellie, and Steven A. Sharpe, 1999, Share repurchases and employee stock options and their implications for S&P 500 share retirements and expected returns, Working paper, Federal Reserve Board.

Liew, Jimmy, and Maria Vassalou, 2000, Can book-to-market, size, and momentum be risk factors that predict economic growth?, *Journal of Financial Economics*, forthcoming.

Lintner, John, 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47, 13–37.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431–468.

Lucas, Robert E., Jr., 1978, Asset prices in an exchange economy, *Econometrica* 46, 1429–1446.

Luttmer, Erzo, 1996, Asset prices in economies with frictions, *Econometrica* 64, 1439–1467.

Lynch, Anthony W., 1999, Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability, Working paper, New York University.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3–28.

Maenhout, Pascal, 1999, Robust portfolio rules and asset pricing, Working paper, Harvard University.

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

Mankiw, N. Gregory, and Jeffrey A. Miron, 1986, The changing behavior of the term structure of interest rates, *Quarterly Journal of Economics* 101, 211–221.

Mankiw, N. Gregory, and Stephen P. Zeldes, 1991, The consumption of stockholders and non-stockholders, *Journal of Financial Economics* 29, 97–112.

Markowitz, Harry, 1952, Portfolio selection, *Journal of Finance* 7, 77–91.

Marsh, Terry, and Robert C. Merton, 1986, Dividend variability and variance bounds tests for the rationality of stock market prices, *American Economic Review* 76, 483–498.

McCulloch, J. Huston, and H. Kwon, 1993, U.S. term structure data, 1947–1991, Working Paper 93–6, Ohio State University.

Mehra, Rajnish, and Edward Prescott, 1985, The equity premium: A puzzle, *Journal of Monetary Economics* 15, 145–161.

Merton, Robert C., 1969, Lifetime portfolio selection under uncertainty: The continuous time case, *Review of Economics and Statistics* 51, 247–257.

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

Merton, Robert C., 1973, An intertemporal capital asset pricing model, *Econometrica* 41, 867–887.

Merton, Robert C., 1980, On estimating the expected return on the market: An exploratory investigation, *Journal of Financial Economics* 8, 323–361.

Modigliani, Franco, and Richard A. Cohn, 1979, Inflation and the stock market, *Financial Analysts Journal* 35, 24–44.

Modigliani, Franco, and Richard Sutch, 1966, Innovations in interest rate policy, *American Economic Review* 56, 178–197.

Nelson, Charles, and M. Kim, 1993, Predictable stock returns: The role of small sample bias, *Journal of Finance* 48, 641–661.

Nelson, William R., 1999, Three essays on the ability of the change in shares outstanding to predict stock returns, Ph.D. dissertation, Yale University.

Odean, Terrance, 1998, Are investors reluctant to realize their losses?, *Journal of Finance* 53, 1775–1798.

Ohlson, J. A., 1990, A synthesis of security valuation theory and the role of dividends, cash flows, and earnings, *Journal of Contemporary Accounting Research* 6, 648–676.

Ohlson, J. A., 1995, Earnings, book values, and dividends in equity valuation, *Journal of Contemporary Accounting Research* 11, 661–687.

Piazzesi, Monika, 1999, An econometric model of the yield curve with macroeconomic jump effects, Working paper, Stanford University.

Poterba, James, and Lawrence H. Summers, 1988, Mean reversion in stock returns: Evidence and implications, *Journal of Financial Economics* 22, 27–60.

Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.

Rietz, Tom, 1988, The equity risk premium: A solution?, *Journal of Monetary Economics* 21, 117–132.

Ritter, Jay R., and Richard S. Warr, 1999, The decline of inflation and the bull market of 1982 to 1997, Working paper, University of Florida.

Roll, Richard, 1977, A critique of the asset pricing theory's tests: Part I, *Journal of Financial Economics* 4, 129–176.

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein, 1985, Persuasive evidence of market inefficiency, *Journal of Portfolio Management* 11, 9–17.

Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.

Ross, Stephen A., 1978, A simple approach to the valuation of risky streams, *Journal of Business* 51, 453–475.

Ross, Stephen A., 1998, Samuelson's fallacy of large numbers revisited, Working paper, MIT.

Samuelson, Paul A., 1965, Proof that properly anticipated prices fluctuate randomly, *Industrial Management Review* 6, 41–49.

Samuelson, Paul A., 1969, Lifetime portfolio selection by dynamic stochastic programming, *Review of Economics and Statistics* 51, 239–246.

Sandroni, A., 1997, Asset prices, wealth distribution, and intertemporal preference shocks, Working paper, University of Pennsylvania.

Santos, M. S., and Michael Woodford, 1997, Rational asset pricing bubbles, *Econometrica* 65, 19–57.

Schroder, Mark, and Costis Skiadas, 1999, Optimal consumption and portfolio selection with stochastic differential utility, *Journal of Economic Theory* 89, 68–126.

Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1153.

Schwert, G. William, 1990, Indexes of U.S. stock prices from 1802 to 1987, *Journal of Business* 63, 399–426.

Shanken, Jay, 1987, Multivariate proxies and asset pricing relations: Living with the Roll critique, *Journal of Financial Economics* 18, 91–110.

Sharpe, Steven A., 1999, Stock prices, expected returns, and inflation, Working paper, Board of Governors of the Federal Reserve System.

Sharpe, William, 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425–442.

Shefrin, Hersh, and Meir Statman, 1985, The disposition to sell winners too early and ride losers too long: Theory and evidence, *Journal of Finance* 40, 777–790.

Shiller, Robert J., 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* 71, 421–436.

Shiller, Robert J., 1982, Consumption, asset markets, and macroeconomic fluctuations, *Carnegie Mellon Conference Series on Public Policy* 17, 203–238.

Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457–498.

Shiller, Robert J., 1994, *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks* (Oxford University Press, Oxford).

Shiller, Robert J., 1999, Human behavior and the efficiency of the financial system, in John Taylor and Michael Woodford, eds.: *Handbook of Macroeconomics Vol. 1* (North-Holland, Amsterdam).

Shiller, Robert J., John Y. Campbell, and Kermit L. Schoenholtz, 1983, Forward rates and future policy: Interpreting the term structure of interest rates, *Brookings Papers on Economic Activity* 1, 173–217.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.

Shleifer, Andrei, 2000, *Inefficient Markets: An Introduction to Behavioral Finance* (Oxford University Press, Oxford).

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.

Siegel, Jeremy, 1999, *Stocks for the Long Run*, 2nd ed. (Irwin).

Stambaugh, Robert F., 1982, On the exclusion of assets from tests of the two parameter model, *Journal of Financial Economics* 10, 235–268.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, 1998, Asset pricing with idiosyncratic risk and overlapping generations, Working paper, Carnegie Mellon University.

Summers, Lawrence H., 1985, On economics and finance, *Journal of Finance* 40, 633–635.

Summers, Lawrence H., 1986, Do we really know that financial markets are efficient? in J. Edwards et al., eds.: *Recent Developments in Corporate Finance* (Cambridge University Press, New York).

Sundaresan, Suresh M., 1989, Intertemporally dependent preferences and the volatility of consumption and wealth, *Review of Financial Studies* 2, 73–88.

Sundaresan, Suresh M., 2000, Continuous-time methods in finance: A review and an assessment, *Journal of Finance*, forthcoming.

Thaler, Richard H., and Eric J. Johnson, 1990, Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice, *Management Science* 36, 643–660.

Tobin, James, 1958, Liquidity preference as behavior towards risk, *Review of Economic Studies* 25, 68–85.

Vasicek, Oldrich, 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177–188.

Vassalou, Maria, 1999, The Fama-French factors as proxies for fundamental economic risks, Working paper, Columbia University.

Vayanos, Dimitri, 1998, Transactions costs and asset prices: A dynamic equilibrium model, *Review of Financial Studies* 11, 1–58.

Veronesi, Pietro, 1999, Stock market overreaction to bad news in good times: A rational expectations equilibrium model, *Review of Financial Studies* 12, 975–1007.

Viceira, Luis M., 1999, Optimal portfolio choice for long-horizon investors with nontradable labor income, NBER Working paper 7409.

Vissing-Jørgensen, Annette, 1997, Limited stock market participation, Working paper, MIT.

Vuolteenaho, Tuomo, 1999, What drives firm-level stock returns? Working paper, University of Chicago.

Wachter, Jessica, 1998, Risk aversion and allocation to long-term bonds, Working paper, Harvard University.

Wachter, Jessica, 1999, Portfolio and consumption decisions under mean-reverting returns: An exact solution for complete markets, Working paper, Harvard University.

Wang, Jiang, 1996, The term structure of interest rates in a pure exchange economy with heterogeneous investors, *Journal of Financial Economics* 41, 75–110.

Weil, Philippe, 1989, The equity premium puzzle and the riskfree rate puzzle, *Journal of Monetary Economics* 24, 401–421.

West, Kenneth D., 1988, Dividend innovations and stock price volatility, *Econometrica* 56, 37–61.

White, Halbert, 2000, A reality check for data snooping, *Econometrica*, forthcoming.

Whitelaw, Robert F., 2000, Stock market risk and return: An equilibrium approach, *Review of Financial Studies*, forthcoming.

Williams, Joseph, 1977, Capital asset prices with heterogeneous beliefs, *Journal of Financial Economics* 5, 219–239.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 1999, Does arbitrage flatten demand curves for stocks? Working paper, Harvard University.

Xia, Yihong, 1999, Learning about predictability: The effect of parameter uncertainty on dynamic asset allocation, Working paper, UCLA.

Xiong, Wei, 1999, Imperfect arbitrage with wealth effects, Working paper, Duke University.