Empirical Asset Pricing: Eugene Fama, Lars Peter Hansen, and Robert Shiller

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:25142544

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Empirical Asset Pricing:
Eugene Fama, Lars Peter Hansen,
and Robert Shiller

John Y. Campbell

May 2014
Abstract

The Nobel Memorial Prize in Economic Sciences for 2013 was awarded to Eugene Fama, Lars Peter Hansen, and Robert Shiller for their contributions to the empirical study of asset pricing. Some observers have found it hard to understand the common elements of the laureates’ research, preferring to highlight areas of disagreement among them. This paper argues that empirical asset pricing is a coherent enterprise, which owes much to the laureates’ seminal contributions, and that important themes in the literature can best be understood by considering the laureates in pairs. Specifically, after summarizing modern asset pricing theory using the stochastic discount factor as an organizing framework, the paper discusses the joint hypothesis problem in tests of market efficiency, which is as much an opportunity as a problem (Fama and Hansen); patterns of short- and long-term predictability in asset returns (Fama and Shiller); and models of deviations from rational expectations (Hansen and Shiller). The paper concludes by reviewing ways in which the laureates have already influenced the practice of finance, and may influence future innovations.

Keywords: Behavioral finance, financial innovation, market efficiency, stochastic discount factor.

JEL classification: G10, G12.
1 Introduction

The 2013 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, awarded for empirical analysis of asset prices, was unforgettably exciting for financial economists. The 2013 laureates, Eugene Fama, Lars Peter Hansen, and Robert Shiller, are giants of finance and architects of the intellectual structure within which all contemporary research in asset pricing is conducted.

The fame of the laureates extends far beyond financial economics. Eugene Fama is one of the world’s most cited economists in any field. Lars Peter Hansen is an immensely distinguished econometrician, so the field of econometrics naturally claims a share of his Nobel glory. Robert Shiller is a founder of behavioral economics, a creator of the Case-Shiller house price indexes, and the author of important and widely read books for a general audience.

The 2013 prize attracted attention in the media, and stimulated discussion among economists, for two additional reasons. First, the behavior of asset prices interests every investor, including every individual saving for retirement, and is a core concern for the financial services industry. Second, the laureates have interpreted asset price movements in strikingly different ways. Robert Shiller is famous for his writings on asset price bubbles, and his public statements that stocks in the late 1990s and houses in the mid 2000s had become overvalued as the result of such bubbles. Eugene Fama is skeptical that the term “bubble” is a well defined or useful one. More broadly, Fama believes that asset price movements can be understood using economic models with rational investors, whereas Shiller does not.

The purpose of this article is to celebrate the 2013 Nobel Memorial Prize in Economic Sciences, and to explain the achievements of the laureates in a way that brings out the connections among them. I hope to be able to communicate the intellectual coherence of the award, notwithstanding the differing views of the laureates on some unsettled questions.

I should say a few words about my own connections with the laureates. Robert Shiller changed my life when he became my PhD dissertation adviser at Yale in the early 1980s. In the course of my career I have written 12 papers with him, the earliest in 1983 and the most recent (hopefully not the last) in 2009. Eugene Fama, the oldest of the 2013 laureates, was already a legend over 30 years ago, and his research on market efficiency was intensively discussed in New Haven and every other center of academic economics. I first met Lars Peter Hansen when I visited Chicago while seeking my first academic job in 1984. I have never forgotten the first conversation I had with him about financial econometrics, in which I sensed his penetrating insight that would require effort to fully understand but would amply reward the undertaking.

Among financial economists, I am not unusual in these feelings of strong connection with the 2013 Nobel laureates. The 2013 award ceremony in Stockholm was notable for the celebratory atmosphere among the coauthors and students of the laureates who were
present, including academics Karl Case, John Cochrane, Kenneth French, John Heaton, Ravi Jagannathan, Jeremy Siegel, and Amir Yaron, central banker Narayana Kocherlakota, and asset management practitioners David Booth, Andrea Frazzini, and Antti Ilmanen. Any of them could write a similar article to this one, although of course the views expressed here are my own and are probably not fully shared by any other economists including the Nobel laureates themselves.

The organization of this paper is as follows. Section 2 gives a basic explanation of the central concept of modern asset pricing theory, the stochastic discount factor or SDF. While the basic theory is not due to the laureates, their work has contributed to our understanding of the concept and many of their empirical contributions can most easily be understood by reference to it.

Section 3 discusses the concept of market efficiency, formulated by Fama in the 1960s. Fama first stated the “joint hypothesis problem” in testing market efficiency, which Hansen later understood to be as much an opportunity as a problem, leading him to develop an important econometric method for estimating and testing economic models: the Generalized Method of Moments.

Section 4 reviews empirical research on the predictability of asset returns in the short and long run. Fama’s early work developed econometric methods, still widely used today, for testing short-run predictability of returns. Typically these methods find very modest predictability, but both Fama and Shiller later discovered that such predictability can cumulate over time to become an important and even the dominant influence on longer-run movements in asset prices. Research in this area continues to be very active, and is distinctive in its tight integration of financial theory with econometrics.

Section 5 discusses the work of the laureates on asset pricing when some or all market participants have beliefs about the future that do not conform to objective reality. Shiller helped to launch the field of behavioral economics, and its most important subfield of behavioral finance, when he challenged the orthodoxy of the early 1980s that economic models must always assume rational expectations by all economic agents. Later, Hansen approached this topic from the very different perspective of robust optimal control.

Section 6 explores the implications of the laureates’ work for the practice of finance. Contemporary methods of portfolio construction owe a great deal to the work of Fama on style portfolios, that is, portfolios of stocks or other assets sorted by characteristics such as value (measures of cheapness that compare accounting valuations to market valuations) or momentum (recent past returns). The quantitative asset management industry uses many ideas from the work of the laureates, and Shiller’s recent work emphasizes the importance of financial innovation for human welfare in modern economies.

Each of these sections refers to the work of more than one of the 2013 Nobel laureates. In this way I hope to foster an appreciation for the intellectual dialogue among the laureates and the many researchers following their lead.
2 The Stochastic Discount Factor: The Framework of Contemporary Finance

2.1 The SDF in complete markets

The modern theory of the SDF originates in the seminal theoretical contributions of Ross (1978) and Harrison and Kreps (1979). Here I present a brief summary in an elementary discrete-state model with two periods, the present and the future, and complete markets.

Consider a simple model with $S$ states of nature $s = 1 \ldots S$, all of which have strictly positive probability $\pi(s)$. I assume that markets are complete, that is, for each state $s$ a contingent claim is available that pays $1$ in state $s$ and nothing in any other state. Write the price of this contingent claim as $q(s)$.

I assume that all contingent claim prices are strictly positive. If this were not true, there would be an arbitrage opportunity in one of two senses. First, if the contingent claim price for some state $s$ were zero, then an investor could buy that contingent claim, paying nothing today, while having some probability of receiving a positive payoff if state $s$ occurs tomorrow, and having no possibility of a negative payoff in any state of the world. Second, if the contingent claim price for state $s$ were negative, then an investor could buy that contingent claim, receiving a positive payoff today, while again having some probability of a positive payoff and no possibility of a negative payoff in the future.

Any asset, whether or not it is a contingent claim, is defined by its state-contingent payoffs $X(s)$ for states $s = 1 \ldots S$. The Law of One Price (LOOP) says that two assets with identical payoffs in every state must have the same price. If this were not true, again there would be an arbitrage opportunity, this time in the sense that an investor could go long the cheap asset and short the expensive one, receiving cash today while having guaranteed zero payoffs in all states in the future. LOOP implies that we must have

$$P(X) = \sum_{s=1}^{S} q(s)X(s). \quad (1)$$

The next step in the analysis is to multiply and divide equation (1) by the objective probability of each state, $\pi(s)$:

$$P(X) = \sum_{s=1}^{S} \frac{\pi(s)}{\pi(s)} q(s)X(s) = \sum_{s=1}^{S} \pi(s)M(s)X(s) = E[MX], \quad (2)$$

where $M(s) = q(s)/\pi(s)$ is the ratio of state price to probability for state $s$, the stochastic discount factor or SDF in state $s$. Since $q(s)$ and $\pi(s)$ are strictly positive for all states $s$, ...
$M(s)$ is also. The last equality in (2) uses the definition of an expectation as a probability-weighted average of a random variable to write the asset price as the expected product of the asset’s payoff and the SDF. This equation is sometimes given the rather grand title of the Fundamental Equation of Asset Pricing.

Consider a riskless asset with payoff $X(s) = 1$ in every state. The price

$$P_f = \sum_{s=1}^{S} q(s) = E[M],$$

(3)

so the riskless interest rate

$$1 + R_f = \frac{1}{P_f} = \frac{1}{E[M]}.$$  

(4)

This tells us that the mean of the stochastic discount factor must be fairly close to one. A riskless real interest rate of 2%, for example, implies a mean stochastic discount factor of $1/1.02 \approx 0.98$.

### 2.1.1 Utility maximization and the SDF

Consider a price-taking investor who chooses initial consumption $C_0$ and consumption in each future state $C(s)$ to maximize time-separable utility of consumption. Assume for now that the investor’s subjective state probabilities coincide with the objective probabilities $\pi(s)$, that is, the investor has rational expectations. The investor’s maximization problem is

$$\text{Max } u(C_0) + \sum_{s=1}^{S} \beta \pi(s) u(C(s))$$

(5)

subject to

$$C_0 + \sum_{s=1}^{S} q(s)C(s) = W_0,$$

(6)

where $W_0$ is initial wealth (including the present value of future income, discounted using the appropriate contingent claims prices). The first-order conditions of the problem can be written as

$$u'(C_0)q(s) = \beta \pi(s) u'(C(s)) \text{ for } s = 1...S.$$  

(7)

These first-order conditions imply that

$$M(s) = \frac{q(s)}{\pi(s)} = \frac{\beta u'(C(s))}{u'(C_0)}.$$  

(8)

In words, the SDF is the discounted ratio of marginal utility tomorrow to marginal utility today. This representation of the SDF is the starting point for the large literature on equilibrium asset pricing, which seeks to relate asset prices to the arguments of consumers’ utility and particularly to their measured consumption of goods and services.
2.1.2 Heterogeneous beliefs

The discussion above assumes that all investors have rational expectations and thus assign the same probabilities to the different states of the world. If this is not the case, we must assign investor-specific subscripts to the probabilities, writing \( \pi_j(s) \) for investor \( j \)’s subjective probability of state \( s \). In general, we must also allow for differences in the utility function across investors, adding a \( j \) subscript to marginal utility as well. Then for any state \( s \) and investor \( j \),

\[
q(s) = \frac{\beta \pi_j(s) u_j'(C_j(s))}{u'_j(C_j0)}.
\]

(9)

The state price is related to the product of the investor’s subjective probability of the state and the investor’s marginal utility in that state. In other words it is a composite “util-prob” to use the terminology of Samuelson (1969).

A similar observation applies to the SDF, the ratio of state price to objective probability:

\[
M(s) = \frac{q(s)}{\pi(s)} = \left( \frac{\pi_j(s)}{\pi(s)} \right) \left( \frac{\beta u'_j(C_j(s))}{u'_j(C_j0)} \right).
\]

(10)

Volatility of the SDF across states may correspond either to volatile deviations of investor \( j \)’s subjective probabilities from objective probabilities, or to volatile marginal utility across states. The usual assumption that investors have homogeneous beliefs rules out the first of these possibilities, while the behavioral finance literature embraces it.

2.1.3 The SDF and risk premia

I now return to the assumption of rational expectations and adapt the notation above to move in the direction of empirical research in finance. I add the subscript \( t \) for the initial date at which the asset’s price is determined, and the subscript \( t + 1 \) for the next period at which the asset’s payoff is realized. This can easily be embedded in a multiperiod model, in which case the payoff is next period’s price plus dividend. I add the subscript \( i \) to denote an asset. Then we have

\[
P_{it} = E_t[M_{t+1}X_{i,t+1}] = E_t[M_{t+1}]E_t[X_{i,t+1}] + \text{Cov}_t(M_{t+1}, X_{i,t+1}),
\]

(11)

where the \( t \) subscripts on the mean and covariance indicate that these are conditional moments calculated using probabilities perceived at time \( t \). The price of the asset at time \( t \) is included in the information set at time \( t \), hence there is no need to take a conditional expectation of this variable. Since the conditional mean of the SDF is the reciprocal of the gross riskless interest rate from (4), equation (11) says that the price of any asset is its expected payoff, discounted at the riskless interest rate, plus a correction for the conditional covariance of the payoff with the SDF.
For assets with positive prices, one can divide through by $P_t$ and use $(1 + R_{i,t+1}) = X_{i,t+1}/P_t$ to get

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

$$= E_t[M_{t+1}]E_t[1 + R_{i,t+1}] + \text{Cov}_t(M_{t+1}, R_{i,t+1}). \quad (12)$$

Rearranging and using the relation between the conditional mean of the SDF and the riskless interest rate,

$$E_t[1 + R_{i,t+1}] = (1 + R_{f,t+1})(1 - \text{Cov}_t(M_{t+1}, R_{i,t+1})). \quad (13)$$

This says that the expected return on any asset is the riskless return times an adjustment factor for the covariance of the return with the SDF.

Subtracting the gross riskless interest rate from both sides, the risk premium on any asset is the gross riskless interest rate times the covariance of the asset’s excess return with the SDF:

$$E_t(R_{i,t+1} - R_{f,t+1}) = -(1 + R_{f,t+1})\text{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1}). \quad (14)$$

### 2.2 Generalizing and applying the SDF framework

The above discussion assumes complete markets, but the SDF framework is just as useful when markets are incomplete. The work of Hansen and Richard (1987) and Hansen and Jagannathan (1991) is particularly important in characterizing the SDF for incomplete markets. Shiller (1982) is an insightful early contribution. Cochrane (2005) offers a textbook treatment.

In incomplete markets, the existence of a strictly positive SDF is guaranteed by the absence of arbitrage—a result sometimes called the Fundamental Theorem of Asset Pricing—but the SDF is no longer unique as it is in complete markets. Intuitively, an SDF can be calculated from the marginal utility of any investor who can trade assets freely, but with incomplete markets each investor can have idiosyncratic variation in his or her marginal utility and hence there are many possible SDFs.

There is however a unique SDF that can be written as a linear combination of asset payoffs and that satisfies the fundamental equation of asset pricing (2). This unique random variable is the projection of any SDF onto the space of asset payoffs, and thus any other SDF must have a higher variance.

#### 2.2.1 Volatility bounds on the SDF

Shiller (1982), a comment by Hansen (1982a), and Hansen and Jagannathan (1991) used this insight to place lower bounds on the volatility of the SDF, based only on the properties of asset returns.
A simple lower bound of this sort can be calculated from a single risky asset return and the return on a riskless asset, using the fact that the correlation between the SDF and any excess return must be greater than minus one. Equation (14) implies that

$$E_t(R_{i,t+1} - R_{f,t+1}) = \frac{-\text{Cov}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1})}{E_t M_{t+1}}$$

$$= \frac{-\text{Corr}_t(M_{t+1}, R_{i,t+1} - R_{f,t+1}) \sigma_t(M_{t+1}) \sigma_t(R_{i,t+1} - R_{f,t+1})}{E_t M_{t+1}}$$

$$\leq \frac{\sigma_t(M_{t+1}) \sigma_t(R_{i,t+1} - R_{f,t+1})}{E_t M_{t+1}}. \tag{15}$$

Rearranging, we get

$$\frac{\sigma_t(M_{t+1})}{E_t M_{t+1}} \geq \frac{E_t(R_{i,t+1} - R_{f,t+1})}{\sigma_t(R_{i,t+1} - R_{f,t+1})}. \tag{16}$$

The standard deviation of the SDF, divided by its mean (which is always close to one given the relation between the mean SDF and the riskless interest rate), must be at least as great as the mean of the risky asset’s excess return divided by its standard deviation, that is, the Sharpe ratio of the risky asset. The tightest lower bound is achieved by finding the risky asset, or portfolio of assets, with the highest Sharpe ratio.

This is a very simple way to understand the famous equity premium puzzle of Mehra and Prescott (1985). If the Sharpe ratio of the aggregate stock market is, say, 0.4, and the average riskless interest rate is, say, 1.03, then the mean SDF must be 1/1.03 or about 0.97, and the standard deviation of the SDF must be at least 0.97 × 0.4 = 0.39. This is a substantial volatility for a random variable with a mean close to one that must always be positive, and it is far greater than the volatilities produced by simple equilibrium models. A consumption-based asset pricing model with a representative agent with power utility, for example, implies that the volatility of the SDF is the coefficient of relative risk aversion times the standard deviation of consumption growth, which is on the order of 0.01, so the volatility bound can only be satisfied for risk aversion coefficients on the order of 40.

Hansen and Jagannathan (1991) generalize the above logic to handle situations where there is no perfectly riskless asset, showing how to calculate a volatility bound for each possible value of a hypothetical riskless interest rate, and hence trace out a minimum-variance frontier for the SDF. They show that this minimum-variance frontier is intimately connected with the minimum-variance frontier for asset returns that plays a key role in the classic mean-variance analysis of Markowitz (1952). Hansen and Jagannathan also show how to tighten the volatility bound by using the fact that the SDF is strictly positive.

The idea of using asset return data to restrict the properties of the SDF remains a fruitful one. More recent work by Stutzer (1995), Bansal and Lehmann (1997), Alvarez and Jermann (2005), and Backus, Chernov, and Zin (2011), for example, shows how asset returns place lower bounds on the entropy of the SDF. Entropy, an alternative to variance as a measure of randomness, is playing an increasingly important role in asset pricing theory as discussed in section 5.3 below.
2.2.2 Conditioning information

The above discussion assumes that economists and investors have the same information set at time $t$ and can use it to calculate conditional moments. Alternatively, one might suppose that investors have more information than is available to economists studying asset returns, a point emphasized by Robert Shiller in his PhD dissertation (1972). Hansen and Richard (1987) is the seminal reference on the effect of conditioning information on tests of asset pricing models.

While Hansen and Richard present a general analysis using advanced methods, their basic point can be understood as follows. One can take unconditional expectations of the conditional asset pricing equation (11), to obtain an unconditional asset pricing formula:

$$
Ep_d = E[M_{t+1}X_{i,t+1}] = E[M_{t+1}]E[X_{i,t+1}] + Cov(M_{t+1}, X_{i,t+1}).
$$

The ability to take unconditional expectations without altering the form of the asset pricing formula is a strength of the SDF approach to asset pricing.

Difficulties arise however when one has an economic model that expresses the SDF as a conditional linear function of some economic variable, for example

$$
M_{t+1} = a_t + b_t R_{m,t+1},
$$

where $R_{m,t+1}$ is the return on the market portfolio as in the classic Capital Asset Pricing Model (CAPM). In this case the conditional covariance $Cov_t(M_{t+1}, X_{i,t+1})$ that appears in equation (11) can be written simply as

$$
Cov_t(M_{t+1}, X_{i,t+1}) = b_t Cov_t(R_{m,t+1}, X_{i,t+1}),
$$

so each asset’s price is risk-adjusted by a time-varying multiple of the conditional covariance of the asset’s payoff with the return on the market. However the unconditional covariance $Cov(M_{t+1}, X_{i,t+1})$ that appears in equation (17) does not take this simple form. Instead,

$$
Cov(M_{t+1}, X_{i,t+1}) = Cov(a_t + b_t R_{m,t+1}, X_{i,t+1}).
$$

As Hansen and Richard emphasize, this implies that even if the CAPM holds conditionally, it need not hold unconditionally. This is also true of other equilibrium asset pricing models such as the consumption CAPM. Hansen and Richard’s analysis has stimulated a subsequent empirical literature. For example, Campbell and Cochrane (1999) and Lettau and Ludvigson (2001) propose conditional consumption-based models to explain failures of unconditional models, while Lewellen and Nagel (2006) and Roussanov (2014) argue that neither the conditional CAPM nor a conditional consumption CAPM can explain the cross-section of stock returns (specifically, the high returns on value stocks relative to growth stocks).
3 The Joint Hypothesis: Problem or Opportunity?

3.1 Market efficiency

Market efficiency is the central concept on which Eugene Fama has built his extraordinary career. Fama (1965) introduced the term, and Fama’s survey “Efficient Capital Markets: A Review of Theory and Empirical Work” (1970) famously defines it by saying that “A market in which prices always ‘fully reflect’ available information is called ‘efficient’.” (p.383).

This definition is evocative rather than precise, as suggested by the use of quotation marks within it. Fama himself admits as much when he writes that this “definitional statement... is so general that it has no empirically testable implications. To make the model testable, the process of price formation must be specified in more detail. In essence we must define somewhat more exactly what is meant by the term ‘fully reflect’.” (p. 384).

In my view the most important contribution of the Fama survey is that it clearly states the so-called joint hypothesis problem, that market efficiency implies a zero conditional mean for asset returns measured relative to the riskless interest rate and the equilibrium compensation for risk. This contribution is made on the second page of the paper, in equation (1), and is worked out more fully in succeeding sections. The implication for empirical researchers is most clearly summarized in Chapter 5 of Fama’s textbook *Foundations of Finance* (1976):

“Some model of market equilibrium, however simple, is required. This is the rub in tests of market efficiency. Any test is simultaneously a test of efficiency and of assumptions about the characteristics of market equilibrium.” (p. 137.)

In the modern language of the stochastic discount factor, Fama’s point can be stated as follows. Suppose we have a model of equilibrium return that determines the right hand side of equation (13), and write this as

\[ Z_{it} = (1 + R_{f,t+1})(1 - \text{Cov}_t(M_{t+1}, R_{i,t+1})). \] (21)

Then equation (13) says that \( Z_{it} \) is the conditional expectation of the return on asset \( i \), so the realized return on the asset satisfies

\[ 1 + R_{i,t+1} = Z_{it} + u_{i,t+1}, \] (22)

where the unexpected return \( u_{i,t+1} \) is unpredictable given any information known at time \( t \). This is testable provided we correctly specify the asset’s covariance with the SDF in calculating \( Z_{it} \), but not otherwise.

To appreciate the importance of this contribution, one can look backward and forward from Fama’s survey. Looking backward, Paul Samuelson (1965) states the Law of Conditional Expectations, that the conditional expectation of a given random variable, to be realized at some date in the future, follows a martingale process. But, while Samuelson does include a
brief discussion of the behavior of a discounted conditional expectation, and he mentions the
case of risk aversion, he does not discuss how his mathematical results can be applied
to models of financial market equilibrium; in fact, he repeatedly disavows his intention to
do so. Another early contribution, Mandelbrot (1966), is similarly disconnected from the
notion of market equilibrium.

Looking forward, Fama’s survey provided a framework for a vast empirical literature
that confronts the joint hypothesis problem and provides a body of meaningful empirical
evidence. Many of the most important early contributions to this literature were made by
Fama himself. To handle the joint hypothesis problem, it was common in the early literature
to assume that expected returns on assets were constant over time, although they could vary
across assets. In other words, the expected return $Z_{it} = Z_i$, a free constant term in a
regression of an asset return on information known at time $t$. The test of market efficiency
was then a test for zero coefficients on all time-varying variables included in the regression.

### 3.1.1 The modern event study

Event studies examine the reaction of asset prices to public news events. Market efficiency
implies that there should be no tendency for systematically positive or negative returns after
news events, except to the extent that the events alter assets’ compensated risk exposures.
If, as traditionally assumed, events have no effect on such risk exposures, the price reaction
at the time of the news event (after controlling for other events occurring at the same time)
is an estimate of the change in fundamental value of the asset (the expected present value
of its dividends, discounted at a constant rate) implied by the news release.

Campbell, Lo, and MacKinlay (1997) date event studies to a Harvard Business Review
study by Dolley (1933). Other event studies appeared in the same publication during the
1950s. However, the methodology that is in use today originates in two papers published in
the late 1960s: Ball and Brown (1968) and Fama, Fisher, Jensen, and Roll (FFJR 1969).
Despite the earlier date of the Ball and Brown paper, it appears to have been written slightly
later and cites FFJR for some of its methodological content. For this reason the Nobel Prize
committee credits FFJR with the invention of the modern event study.

Key contributions of the FFJR paper include the use of a market model regression to
adjust price movements for contemporaneous movements in the aggregate stock market, and
the averaging of price reactions across separate events through the graphical device of an
event-time diagram. The event-time diagram makes it easy to see the effect of information
leakage before an event, and to evaluate the prediction of market efficiency that prices have
no tendency to drift upwards or downwards after the event relative to their equilibrium path
(which will have a negligible tilt over short periods of a few days or weeks).

The FFJR paper does not contain any statistical analysis—there are no confidence inter-
vals in its event-time diagrams—but the insight of the paper, that noise can be reduced by
averaging across events that take place at different calendar times, since unexpected returns are uncorrelated across calendar time under market efficiency, is the basis of the modern statistical theory of event studies (see for example the textbook treatment in Chapter 4 of Campbell, Lo, and MacKinlay 1997).

The enormous empirical literature using event studies finds that many events have immediate impacts with no subsequent tendency for drift in prices. Some events, however, do seem to be followed by drifts in prices in the same direction. The most famous example is earnings announcements (Ball and Brown 1968), although post-earnings-announcement drift has weakened in recent years, perhaps because arbitrageurs have exploited the phenomenon. These academic findings have had considerable influence on the legal system (Gilson and Kraakman 1984). Securities and corporate litigators routinely rely on event studies to infer fundamental information from price movements.

3.1.2 Firm characteristics as stock return predictors

Data on the characteristics and returns of common stocks form a panel, and empirical models predicting returns from stock characteristics can be estimated using panel regression methods. Unlike standard microeconomic panels, however, a panel of stock returns has strong cross-sectional correlation arising from common shocks that move groups of stock returns together. For example, high-beta stocks all tend to outperform the market when the market goes up; and stocks in a particular industry tend to move together even controlling for market movements.

The simple but powerful insight of Fama and MacBeth (1973) is that market efficiency, with constant expected returns over time, implies that stock returns are uncorrelated over time even though they are correlated across stocks at a given time. In this sense a finance panel regression has a structure that is orthogonal to a standard microeconomic panel regression (which has time-series correlation of variables for a given household, but assumes no correlation of variables across households). In modern panel regression terminology, the finance panel should cluster standard errors by time while a microeconomic panel should cluster standard errors by household.

Fama and MacBeth present a brilliantly simple way of doing this. They suggest estimating a sequence of cross-sectional regressions of stock returns on characteristics. In each cross-section, the coefficients can be interpreted as returns on portfolios of stocks weighted by firm characteristics (an interpretation developed in Fama 1976, Chapter 9). The returns on these portfolios are serially uncorrelated given market efficiency and the assumption that the included characteristics include all those that determine expected returns. Then the time-series average return on each portfolio estimates the average effect of the given characteristic (controlling for other included characteristics), and its standard error can be calculated from the time-series variability of the portfolio return.
The account of the Fama-MacBeth method that I have just given interprets the method as a way to estimate how observable firm characteristics affect average stock returns. This is not the way Fama and MacBeth themselves present the method. They are interested in testing the Capital Asset Pricing Model (CAPM): either the original model of Sharpe (1964) and Lintner (1965), or a more general version with an unrestricted intercept, as in Black (1972). The CAPM relates average returns to firms’ betas with the aggregate stock market, which are themselves unobservable parameters that must be estimated. Fama and MacBeth, following Black, Jensen, and Scholes (1972), argue that estimation error in betas can be reduced by first sorting stocks by their beta estimates in lagged data, then forming portfolios of stocks with similar past beta estimates and using these portfolios as test assets. The Fama-MacBeth method continues to be used as a convenient way to test the CAPM against specific alternatives, but subsequent tests of the CAPM often use the methodology of Gibbons, Ross, and Shanken (1989) which avoids treating betas as known parameters.

3.2 From problem to opportunity

In the late 1970s, economists began to understand that the joint hypothesis problem could also be an opportunity. If a model of market equilibrium, together with the hypothesis of rational expectations, restricts the data by implying that abnormal returns are unpredictable, then the unpredictability of abnormal returns can be used to estimate the unknown parameters of the model. Furthermore, the notion of unpredictability applies not only to returns, but also to the marginal utility of investors given the role played by marginal utility in the stochastic discount factor. This implies restrictions on the joint dynamics of asset returns and the arguments of investors’ utility, most obviously their consumption of goods that provide an immediate utility flow.

Robert Hall was the first to realize that an economic model implies unpredictability of changes in consumption. Hall (1978) shows that under certain restrictive assumptions, current consumption is the best predictor of future consumption, and tests this by regressing future consumption on current consumption and other candidate predictors, looking for coefficients of one and zero respectively. Hall’s paper cites Fama’s efficient markets survey, although not his papers on predictive regression.

3.2.1 Generalized method of moments

Lars Peter Hansen exploited these ideas in a much more systematic way. The Generalized Method of Moments (GMM) of Hansen (1982b) is an econometric approach that is particularly well suited for testing models of the SDF. As Hansen (2001) reviews, GMM builds on a statistics literature going back to Karl Pearson’s late-19th Century Method of Moments and the mid-20th Century contributions of Neyman (1949) and Sargan (1958, 1959).
To understand GMM, begin by writing the Fundamental Equation of Asset Pricing, equation (2), as
\[ P_t = E_t[M_{t+1}(b)X_{t+1}], \]  
(23)
where \( b \) is a vector of parameters. Next take unconditional expectations to get
\[ E[P_t] = E[M_{t+1}(b)X_{t+1}] \]  
(24)
or
\[ E[M_{t+1}(b)X_{t+1} - P_t] = E[u_{t+1}(b)] = 0. \]  
(25)
Here \( u_{t+1}(b) \) is the pricing error of the model, which depends on the parameters of the model for the SDF. If we have a vector of asset returns, then \( u_{t+1}(b) \) is a vector.

There are several variants of this idea. For example, we might divide through by prices, expressing the model in terms of returns:
\[ E[M_{t+1}(b)(1 + R_{t+1}) - 1] = 0. \]  
(26)
Or we might premultiply the conditional equation by instruments \( Z_t \) known at time \( t \), then take unconditional expectations to get
\[ E[Z_tM_{t+1}(b)(1 + R_{t+1}) - Z_t] = 0. \]  
(27)
In all these cases we will write \( u_{t+1}(b) \) for the pricing error, the object that should have unconditional expectation equal to zero if the SDF model is correctly specified and the true parameter vector is used to calculate the pricing error. With \( N \) asset returns and \( J \) instruments, \( u_{t+1}(b) \) is an \( NJ \times 1 \) vector.

We define \( g_T(b) \) as the sample mean of the \( u_{t+1}(b) \) in a sample of size \( T \):
\[ g_T(b) = \frac{1}{T} \sum_{t=1}^{T} u_{t+1}(b). \]  
(28)

The first-stage GMM estimate of \( b \) solves
\[ \min g_T(b)'Wg_T(b) \]  
(29)
for some weighting matrix \( W \). This estimate, \( \hat{b}_1 \), is consistent and asymptotically normal. Formulas for the asymptotic variance-covariance matrix of \( \hat{b}_1 \) and the minimized first-stage objective function enable asymptotic inference as explained in Cochrane (2005).

We can continue to a second stage. Using \( \hat{b}_1 \) we form an estimate \( S_T(\hat{b}_1) \) of
\[ S = \sum_{j=-\infty}^{\infty} E[u_{t+1}(b)u_{t+1-j}(b)'], \]  
(30)
the long-run variance-covariance matrix of the pricing error. Standard methods, such as those of Hansen and Hodrick (1980) or Newey and West (1987), can be used to do this consistently even in the presence of arbitrary heteroskedasticity and autocorrelation in the pricing error.

Then the second-stage GMM estimate of $b$ solves

$$\min g_T(b)' S_T(\hat{b}_1)^{-1} g_T(b).$$

(31)

This estimate, $\hat{b}_2$, is consistent, asymptotically normal, and asymptotically efficient (that is, it has the smallest variance-covariance matrix among all choices of weighting matrix $W$).

The variance-covariance matrix of $\hat{b}_2$ is given by

$$\text{Var}(\hat{b}_2) = \frac{1}{T} (d' S^{-1} d)^{-1},$$

(32)

where

$$d = \frac{\partial g_T(b)}{\partial b},$$

(33)

and we can estimate it consistently using sample estimates of $S$ and $d$. Also, the model can be tested using the following result for the asymptotic distribution of the minimized second-stage objective function:

$$T \min g_T(b)' S_T(\hat{b}_1)^{-1} g_T(b) \sim \chi^2(NJ - K),$$

(34)

where $NJ$ is the number of moment conditions and $K$ is the number of parameters.

This framework is extremely general. The usual formulas for OLS and GLS regression can be derived as special cases. The decision whether to proceed to a second-stage GMM estimate is equivalent to the decision whether to use GLS. The second-stage estimate is more efficient asymptotically, but can behave poorly in finite samples if the matrix $S$ is poorly estimated. This would be the case, for example, if the number of moment conditions is large relative to the sample size.

Hansen, Heaton, and Yaron (1996) have proposed an alternative to two-stage GMM, continuously updated GMM, that uses a different weighting matrix for every parameter vector considered, in other words solving

$$\min g_T(b)' S_T(b)^{-1} g_T(b).$$

(35)

This appears to have better finite-sample properties than does two-stage GMM (Newey and Smith 2004), and is increasingly used in recent empirical research.

Other empirical researchers favor single-stage GMM on the grounds that first-stage GMM estimates with a sensible weighting matrix can be more persuasive than more sophisticated two-stage GMM or maximum likelihood estimates, just as OLS or WLS regression estimates
with exogenously specified weights can be more persuasive than GLS estimates. One advantage of this approach is that the minimized objective function varies only with the elements of the pricing error vector, because the matrix used to weight them is held constant. This makes it easier to compare the objective function across alternative model specifications.

In all versions of GMM, but particularly in a single-stage approach, the initial weighting matrix can influence the results and should be chosen thoughtfully. One simple choice is the identity matrix. Another, advocated by Hansen and Jagannathan (1997), is the inverse of the variance-covariance matrix of asset returns. This has some of the advantages of the optimal weighting matrix but can be calculated without reference to first-stage parameter estimates.

GMM is an important econometric method for several reasons. It enables econometricians to test economic models without having to take a stand on secondary features of the economic environment, in other words, without having to make auxiliary assumptions that would be needed to write down and maximize a likelihood function. GMM also avoids distributional assumptions on economic shocks, and makes it straightforward to handle nonlinear models. Many econometricians feel that GMM is a contribution that could have been recognized with a Nobel prize independent of its connection with empirical asset pricing. While I do not disagree with this assessment, it is no coincidence that the first applications of GMM by Hansen and Singleton (1982, 1983) and other early applications such as Eichenbaum, Hansen, and Singleton (1988) were to asset pricing models. The immense popularity of GMM among both financial economists and macroeconomists is due in part to its alignment with the theory of efficient markets developed by Fama.

4 Predicting Asset Returns in the Short and Long Run

4.1 The information in asset prices

Another important strand of Fama’s research uses predictive regressions to extract information in asset prices. The earliest paper of this type is “Short-Term Interest Rates as Predictors of Inflation” (Fama 1975). The idea is extremely simple. Given an auxiliary asset pricing model (in this case, an assumption that the equilibrium real interest rate is constant), the market’s forecast of future inflation can be extracted from the nominal interest rate, up to an unknown constant (the value of the real interest rate). If the market is efficient, then this forecast contains all publicly available information relevant for predicting future inflation. A regression of future inflation on a constant and the nominal interest rate should have a coefficient of one on the nominal interest rate, and if additional variables such as lagged inflation rates are added to the regression, the coefficients on these variables should all be zero. Fama found that US data from 1953–71 were consistent with these predictions, although the result broke down shortly afterwards as the real interest rate became highly
variable in the 1970s and 1980s.

The importance of Fama (1975) lies not in its conclusion, but in its method. The paper tests a model of interest-rate determination by using the interest rate not as the dependent variable in a regression, but as an explanatory variable. This was counterintuitive at the time (although logical given the theory of efficient markets), but it seems natural today because of the many papers that have followed its lead. Some of these are by Fama himself, including important papers on the interest differential across currencies (Fama 1984a), the term structure of Treasury bill rates (Fama 1984b), and the term structure of Treasury bond yields (Fama and Bliss 1987).

These Fama papers documented predictable time-variation in excess returns across currencies and bonds of different maturities, providing some of the first evidence against simple models that make risk premia constant over time. The interest differential between two currencies, for example, should predict depreciation of the high-interest-rate currency if expected rates of return are equal in the two currencies; in fact, it predicts appreciation of the high-interest-rate currency, implying large excess returns in that currency, a phenomenon that is the basis for the currency carry trade. Similarly, the yield spread between two interest rates of different maturities should predict increases in both short-term and long-term interest rates if expected returns are equal across maturities; in fact, as claimed by Macaulay (1936) and shown by Shiller, Campbell, and Schoenholtz (1983) and Campbell and Shiller (1991) as well as Fama and Bliss (1987), long-term rates tend to decline after yield spreads become unusually wide.

4.2 Variance bounds and long-run return predictability

By the late 1970s, a body of research had tested simple asset pricing models with constant expected returns, or expected excess returns relative to the riskfree interest rate, and had found only minor deviations from such models. (Most of this research looked at equity markets. The work of Fama and others on currencies and fixed-income securities still lay in the future.) At this time Michael Jensen (1978) felt confident enough to write “There is no other proposition in economics which has more solid evidence supporting it than the Efficient Markets Hypothesis”.

This was the moment at which Shiller (1981) launched an important critique of the prevailing orthodoxy. Shiller’s approach is based on the following observation. Return to equation (11) and rewrite it as

$$P_t = E_t[M_{t+1}X_{t,t+1}] = \left( E_t[M_{t+1}] + \frac{Cov_t(M_{t+1}, X_{t,t+1})}{E_t[X_{t,t+1}]} \right) E_t[X_{t,t+1}] = \delta_t E_t[X_{t,t+1}].$$

If the discount rate $\delta$ is constant over time at a level $\delta$, as assumed by the early literature on market efficiency, and if the payoff next period is the sum of price plus dividend next
period, \( X_{i,t+1} = P_{i,t+1} + D_{i,t+1} \), then this difference equation can be solved forward. Under the assumption that future prices are not explosive, the discounted future price eventually becomes negligible and we obtain the dividend discount model for the asset price:

\[
P_{it} = E_t \sum_{j=1}^{\infty} \delta^j D_{i,t+j},
\]

(37)

Future expectations drop out of this expression because the Law of Iterated Expectations tells us that the expectation today of future expectations of future dividends is the same as the expectation today of future dividends.

Shiller (1981) observed that if (37) holds, the realized discounted value of future dividends should equal the stock price plus unpredictable noise, and therefore should have greater variance than the stock price. He calculated a proxy for realized discounted dividends on an aggregate stock index, using a terminal condition to account for dividends not yet paid, and found that this series has much lower variance than the price of the index, contrary to the prediction of the model. LeRoy and Porter (1979) made a similar observation, and Shiller (1979) conducted a related analysis of long-term bond yields.

Shiller’s critique generated a major controversy. Kleidon (1986) and Marsh and Merton (1986) emphasized that both dividends and stock prices follow highly persistent processes with unit roots, in which case the population variances of prices and of realized discounted dividends are undefined. Sample variances can be calculated in any sample, but they increase without limit as the sample size increases.

In response to this, Campbell and Shiller (1987, 1988a, 1988b) showed how to modify the variance calculations for the unit root case. A linear model for dividends, in which the dividend process has a unit root, implies that prices and dividends are cointegrated with a cointegrating parameter that depends on the discount rate \( \delta \) in (37). Campbell and Shiller (1987) tested and rejected this form of the dividend discount model, once again finding excessive volatility—this time in the spread between prices and current dividends rather than in the level of prices.

While linear present value models are tractable, loglinear models are more appealing empirically. Campbell and Shiller (1988a) introduced a loglinear approximate framework that has been used in much subsequent empirical research. By taking a Taylor approximation of the nonlinear equation relating log returns to log prices and log dividends, around the mean of the log dividend-price ratio, and solving forward the resulting loglinear difference equation, Campbell and Shiller found that

\[
p_t \approx \frac{k}{1 - \rho} + \sum_{j=0}^{\infty} \rho^j [(1 - \rho)d_{t+1+j} - r_{t+1+j}],
\]

(38)

where lower-case letters denote logs, \( k \) and \( \rho \) are parameters of loglinearization, and the asset-specific subscript \( i \) has been dropped for notational simplicity.
This approximate equation holds ex post, as an accounting identity. It should therefore hold ex ante, not only for rational expectations but for any expectations that satisfy identities. Since the stock price at time $t$ is known at time $t$, it follows that the stock price can be written as a discounted sum of expected future log dividends and returns. These two components can be thought of as the “cash-flow” and “discount-rate” components of the price.

If log dividends follow a unit root process, then log dividends and log prices are cointegrated with a known cointegrating vector. The log dividend-price ratio is stationary, and is approximately linearly related to dividend growth and returns:

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

(39)

Again, this equation must hold ex ante as well as ex post, so one can decompose the variance of the log dividend-price ratio into components related to expected future dividend growth (cash flows) and time-varying discount rates. Campbell and Shiller (1988a, 1988b) did this using vector autoregressions forecasting returns (or dividend growth) with other variables including the log dividend-price ratio. Since they calculated expected returns from an econometric forecasting model, they were estimating the discount rates that would be applied to cash flows by an investor with rational expectations.

The late 1980s saw a convergence between the apparently very different research agendas and worldviews of Fama and Shiller. The methods developed by Campbell and Shiller calculated the contribution of time-varying discount rates to the volatility of the log dividend-price ratio, using VAR forecasts of long-run discounted stock returns. Fama and French (1988a) ran direct regressions of long-horizon returns onto the dividend-price ratio, correcting standard errors for overlap in long-horizon returns using methods developed earlier by Hansen and Hodrick (1980) and Newey and West (1987). They found high explanatory power for these regressions. Fama and French (1988b) and Poterba and Summers (1988), in related work, reported evidence for negative serial correlation of stock returns at annual and lower frequencies. All these results implied that time-varying discount rates—that is, rational expectations of future returns—are important for understanding the variability of the dividend-price ratio in the aggregate stock market. While Fama and Shiller disagree about the interpretation of these findings, they do not disagree about the facts.

4.3 Finance theory and return predictability

The literature on return predictability has remained active over the past 25 years. Variance decompositions can be calculated not only for the stock price (in the case where this is stationary), the spread between prices and dividends (when the dividend follows a linear process with a unit root), and the log dividend-price ratio, but also for other valuation ratios such as the log ratio of prices to smoothed earnings (Campbell and Shiller 1988b) and the
log ratio of enterprise value to total payout (Larrain and Yogo 2008), or even for the log ratio of consumption to total wealth (Lettau and Ludvigson 2001a). Each of these ratios may have a different decomposition, since transitory variation in a payout measure implies variability in expected future payout growth, which in turn contributes to the volatility of the corresponding valuation ratio. As an alternative approach, Campbell (1991) suggested calculating variance decompositions for returns rather than for valuation ratios. Variance decompositions can also be calculated for individual stocks and style portfolios, although this requires confronting nonstationarities in firm-level payout policy that are less important at the aggregate level (Vuolteenaho 2002, Cohen, Polk, and Vuolteenaho 2009).

There has been considerable concern about the small-sample properties of regressions predicting stock returns from valuation ratios. Stambaugh (1999) pointed out that when the explanatory variables in return-predicting regressions are persistent and have innovations that are correlated with returns (as is certainly the case for valuation ratios), the coefficients are biased upwards. A similar problem afflicts t-statistics as shown by Cavanagh, Elliott, and Stock (1995).

During the 2000s, a number of papers proposed to use theoretical restrictions to improve the power of tests of return predictability. Lewellen (2004) showed that when theory tells us that the log dividend-price ratio cannot be explosive, it is possible to compute a test statistic under the worst-case assumption that this ratio has a unit root. In samples where the valuation ratio appears to have a root very close to unity, Lewellen’s test can reject more strongly than the standard test — and this is exactly what happens in an application to US data. Campbell and Yogo (2006) propose a related procedure.

Cochrane (2008) emphasizes the inability of the log dividend-price ratio to predict dividend growth. Using the Campbell-Shiller approximation, he notes that if the dividend-price ratio fails to predict stock returns positively, it will be explosive unless it predicts dividend growth negatively. Appealing like Lewellen (2004) to a theoretical presumption that the dividend-price ratio cannot be explosive, Cochrane argues that the absence of predictable dividend growth strengthens the evidence for predictable returns.

Several other recent papers also use restrictions from finance theory to improve forecasts of stock returns. Goyal and Welch (2008) direct attention to the poor out-of-sample performance of return predictions based on regressions that include both a constant and time-varying explanatory variables such as the dividend-price ratio. The difficulty with these regressions is partly that they must estimate the unconditional mean of stock returns using noisy historical data. In response, Campbell and Thompson (2008) show that out-of-sample performance is improved by imposing theoretically motivated sign restrictions on the regression coefficient and the fitted value, and improved further by using a version of a Gordon growth model to avoid directly estimating the unconditional mean of stock returns. Likewise, Fama and French (2002) argue that the unconditional mean stock return can be better estimated by correcting historical average returns for the historical change in valuation ratios, because finance theory implies that such changes cannot be expected
to continue. Avdis and Wachter (2013) make a similar correction in a formal maximum likelihood framework.

A theme of this econometric literature is that finance theory gives us extra information about stock price valuation that can be used to improve the obvious procedure of predicting returns by standard OLS regressions on information variables. In the words of Shiller (2013), “testing market efficiency by regressing excess returns on information variables makes no use of the terminal condition that requires that all movements in prices need to be justified by information about subsequent movements in fundamentals”.

During the same quarter-century, finance theorists have built increasingly sophisticated equilibrium models that are consistent with the econometric evidence for return predictability. The focus has been on mechanisms that generate time-variation in risk premia, often related to the business cycle given the evidence for cyclical variation in risk premia presented in Fama and French (1989), and sometimes capturing acyclical variation that has been highlighted in more recent research.

One class of models, following Merton (1980), has time-varying volatility of aggregate stock returns and an equity premium that moves in proportion to either the standard deviation of returns (implying a constant Sharpe ratio) or the variance of returns (implying a constant allocation to stocks in a static portfolio choice model). A vast empirical literature shows that stock market volatility moves over time, either estimating parametric models such as ARCH models (Engle 1982, Bollerslev 1986), or estimating stochastic volatility with non-parametric methods based on high-frequency data (Barndorff-Nielsen and Shephard 2002), or extracting implied volatility from option prices. However, while there is some evidence the equity premium is positively related to equity volatility, it does not seem to move in proportion with either the standard deviation or variance of aggregate stock returns (Campbell 1987, French, Schwert, and Stambaugh 1987, Ghysels, Santa-Clara, and Valkanov 2005, Harvey 1989). Hence the literature has sought to model time-variation in the reward that investors require for bearing equity risk.

Because of the relation discussed above between the maximum Sharpe ratio and the volatility of the stochastic discount factor, models of time-variation in the equity Sharpe ratio require time-variation in the volatility of the SDF. This can be achieved, while maintaining the assumption that investors have rational expectations, through time-variation in the volatility of aggregate consumption growth (as in the long-run risk literature following Kandel and Stambaugh 1991, Bansal and Yaron 2004, Hansen, Heaton, and Li 2008, and Hansen 2012, or in the model of Martin 2013), tail risk in the consumption distribution (as in Gabaix 2012 and Wachter 2013), curvature of the utility function (as in the habit formation model of Campbell and Cochrane 1999), or uninsurable idiosyncratic risk (as in Constantinides and Duffie 1996, Mankiw 1986, and Storesletten, Telmer, and Yaron 2008).

During the last decade this literature has expanded further, to confront not only the predictability of aggregate returns but also the variation in average returns within the cross-
section of equities. Often the focus is on the average returns of the value- or size-sorted stock portfolios introduced by Fama and French (1993). A complete review would take me too far afield, but the contributions of Berk, Green, and Naik (1999), Campbell and Vuolteenaho (2004), Hansen, Heaton, and Li (2008), Lettau and Wachter (2007), Parker and Julliard (2005), and Yogo (2006) are illustrative of this literature.

This model development can be regarded cynically, as evidence that the joint hypothesis problem makes market efficiency ultimately untestable. I regard it more hopefully, as a sign that financial economics is a mature science in which evidence guides the development of theory whose novel predictions can then be tested in a process that moves us gradually towards a deeper understanding of asset markets.

5 Beyond Rational Expectations: Behavioral Finance and Ambiguity Aversion

5.1 Beyond rational expectations

The theory of rational expectations, first stated by Muth (1961), was used in the late 1960s and the 1970s to transform macroeconomics. By the early 1980s the pioneering rational expectations macroeconomists, led by Robert Lucas and Thomas Sargent, had developed a set of macroeconomic models and accompanying econometric methods that assumed rational expectations and used the restrictions implied by this assumption to estimate unknown model parameters and test the models. This work proceeded in parallel with the development of efficient markets theory and econometrics led by Fama and Hansen.

In retrospect, the early 1980s were the high tide for rational expectations theory. As the tide turned, a number of economists, including first Shiller and later Hansen and Sargent, began to look for alternatives to the assumption that economic agents’ subjective expectations always equal objective expectations.

Asset pricing is a natural context in which to consider irrational expectations, because the theory of the stochastic discount factor tells us exactly how to find subjective expectations that can rationalize asset prices without any risk aversion on the part of investors. These so-called risk-neutral expectations can be found as follows. Starting from the relation between asset prices and state prices, \( P(X) = \sum q(s)X(s) \), multiply and divide by the sum of the state prices, \( \sum q(s) = P_f = 1/(1 + R_f) \), to obtain

\[
P(X) = \frac{1}{1 + R_f} \sum_{s=1}^{S} \pi^*(s)X(s) = \frac{1}{1 + R_f} E^*[X].
\]

(40)

Here \( \pi^*(s) = q(s) / \sum q(s) \) is a “pseudo-probability” of state \( s \). It has the necessary proper-
ties to be a probability, namely that it is positive and sums to one across all possible states. However it is derived only from state prices and need not correspond to the objective probability of state $s$. Equation (40) says that the asset price equals the pseudo-probability-weighted average payoff, or equivalently the pseudo-expectation of the payoff, discounted at the riskless interest rate. For this reason $\pi^*(s)$ is also known as a risk-neutral probability.

If all investors have subjective probabilities $\pi^*(s)$, then observed asset prices can be reconciled with risk-neutral preferences. Put another way, any asset pricing model that relies on utility curvature and rational expectations can be matched by a model with linear utility and particular irrational expectations. A similar procedure can be used to reconcile asset prices with an arbitrary risk-averse utility function. Thus asset price data alone cannot tell us whether investors have rational expectations; we need some external evidence on either preferences or expectations to resolve the issue. This is of course another way to state Fama’s joint hypothesis problem, and it helps to explain how Fama and Shiller can simultaneously agree about empirical patterns in asset prices and disagree about the interpretation of these patterns.

5.2 Behavioral finance

During the 1980s Robert Shiller moved beyond a critique of rational expectations models with constant discount rates to articulate an alternative vision of financial market equilibrium. His Brookings paper “Stock Prices and Social Dynamics” (1984) is an important first step in this direction. It is striking not only for its content but also for its style, which is far more literary than is typical for an economics paper and which prefigures Shiller’s later writing for a general audience in books such as *Irrational Exuberance* (2000).

The first part of the 1984 paper argues that the true model of the economy is unknowable to economists and investors alike. In this context, subjective views about future economic prospects and asset payoffs spread among investors in a manner analogous to the spread of infections in epidemiological models. Thus we should expect asset prices to be influenced by “fashion... the great governor of this world” in the memorable phrase of Henry Fielding.

The second part of the paper writes down a simple model of equilibrium in a stock market with both rational and irrational investors. The equity demand function of rational investors is linear in the expected return, as would be implied by a utility function with constant absolute risk aversion in an environment with normally distributed risks. Irrational investors demand an exogenous value of shares; equivalently, their equity demand function has unit price elasticity. In this model stock prices can be written as a discounted present value of future dividends and exogenous irrational-investor demands. Irrational investors have a larger effect on stock prices when they are more persistent, and when the risk-bearing capacity of rational investors is smaller.
Shiller (1984) was an early contribution to what has become a vast literature on behavioral finance. This literature addresses a number of important questions about equilibrium with both rational and irrational investors. First, what determines the asset demands of irrational investors? Shiller (1984) modeled these demands as an exogenous stochastic process, but it is appealing to derive them from assumptions about irrational investors’ expectations and perhaps also their preferences. For example Barberis, Shleifer, and Vishny (1998) model expectations as switching between two false models in a way that generates both short-run momentum and longer-run mean reversion, while Cecchetti, Lam, and Mark (2000) model irrational expectations about the persistence of booms and recessions. Other behavioral research has emphasized overconfidence in private information (Daniel, Hirshleifer, and Subrahmanyam 1998, 2001), direct utility from optimistic anticipation of the future (Bénabou 2013 and Brunnermeier and Parker 2005), and time-varying risk aversion driven by nonstandard prospect-theory preferences (Barberis, Huang, and Santos 2001, Benartzi and Thaler 1995, Kahneman and Tversky 1979, Thaler and Johnson 1990). Shiller (1988) and Shiller and Pound (1989) used surveys around the time of the 1987 stock market crash to understand investor belief formation, and survey data have also been used by Barberis, Greenwood, Jin, and Shleifer (2013), Frankel and Froot (1989), Froot (1989), and Greenwood and Shleifer (2013) among others.

Second, what prevents rational investors at a point in time from arbitraging away the effects of irrational investors on asset prices? The most obvious answer, and the one discussed by Shiller (1984), is that rational investors are risk-averse. As rational investors trade with irrational investors, they take on more or less stock market exposure and the covariance of their marginal utility with stock returns varies accordingly, justifying a time-varying expected stock return. During a boom, when irrational investors are optimistic, rational investors hold fewer stocks, their marginal utility comoves less with stock returns, and an SDF derived from rational investors’ marginal utility implies that the expected stock return declines. In the extreme, if rational investors actually short stocks in order to supply the demand of irrational investors, their marginal utility may comove negatively with stock returns in which case the equilibrium stock return becomes negative. Conversely, rational investors’ stock exposures increase during a downturn, justifying a higher expected stock return at such a time. This logic implies that even within a behavioral model, the risk assessments of rational investors remain relevant.

The behavioral literature explores other answers to this question. For example, there may be short-sales constraints so that pessimistic rational investors cannot offset the demands of optimistic irrational investors (Harrison and Kreps 1978, Miller 1977, Scheinkman and Xiong 2003). Also, rational investors may be financial intermediaries whose clients pull their capital after losses have been incurred (Brunnermeier and Pedersen 2009, Shleifer and Vishny 1997).

Third, why don’t rational investors become richer than irrational investors over the long run, ultimately minimizing the price impacts of irrational investors as conjectured by Friedman (1953)? Shiller (1984) answers this objection by pointing out that wealthy rational
investors eventually die and leave their money to less rational descendants. DeLong, Shleifer, Summers, and Waldmann (1990a,b) argue that rational investors may be more risk-averse than irrational investors, whose willingness to earn a risk premium may outweigh their poor market timing and allow them to accumulate wealth. Kogan, Ross, Wang, and Westerfield (2006) show that irrational investors can have a significant impact on asset prices even when their wealth is small relative to that of rational investors.

In parallel with the rational asset pricing literature, the behavioral finance literature has also explored asset pricing patterns within the cross-section of asset returns. There is a particularly active debate over the rationality or otherwise of the value premium, the anomalously high returns to value stocks, and the related phenomenon of long-run mean-reversion in individual stock returns highlighted by De Bondt and Thaler (1985). Momentum, the tendency of high returns over the past year (excluding the last 1–3 months) to predict high future returns, is also a favorite target for behavioral modeling and is much more difficult for rational models to explain. Selected papers from this literature include Baker and Wurgler (2006), Chan, Jegadeesh, and Lakonishok (1996), Hong, Lim, and Stein (2000), Hong and Stein (1999), and LaPorta, Lakonishok, Shleifer, and Vishny (1997).

Behavioral finance is part of a much broader development of behavioral economics that has occurred since the early 1980s. Finance is one of the most successful behavioral fields, both because data are increasingly available on individual financial decisions, and because behavioral finance has models of equilibrium with both rational and irrational investors that are simple enough to be solved for their pricing implications. Robert Shiller and Richard Thaler, who together direct the working group on behavioral economics at the National Bureau of Economic Research, are widely regarded as pioneers not only of behavioral finance, but also of behavioral economics more generally.

5.3 Ambiguity aversion

The view that investors do not know the true model of the economy, which motivated Shiller’s (1984) development of behavioral finance, is also the starting point for a theoretical literature on ambiguity aversion. Building on the insights of Knight (1921) and the experimental evidence of Ellsberg (1961), this literature argues that investors handle uncertainty about models differently from uncertainty about outcomes within a model. Within a model, outcomes can be described by a probability distribution, but investors do not behave as if they have a subjective probability distribution over alternative models and therefore cannot be described as Bayesians in their response to model uncertainty or “ambiguity”. Instead, authors such as Gilboa and Schmeidler (1989) and Klibanoff, Marinacci, and Mukerji (2005) have argued that investors behave conservatively with respect to ambiguity, acting as if a worst-case model is true. Epstein and Schneider (2010) offer a recent review.
In a series of papers and books written with Thomas Sargent and other coauthors during the last twenty years (notably Anderson, Hansen, and Sargent 2003, Hansen and Sargent 1995, 2001, and 2008 and Cagetti, Hansen, Sargent, and Williams 2002), Lars Peter Hansen has adapted an engineering literature on robust optimal control to develop one of the major paradigms of this literature. His work in this area flows naturally from his concern, as an econometrician, with model misspecification and his earlier work with Ravi Jagannathan (1997) on misspecified models of the SDF.

In the Hansen-Sargent framework, an investor has a reference model of the economy but chooses the actions that maximize expected utility in a “worst-reasonable-case” model that delivers the lowest expected utility within a set of models that do not deviate too far from the reference model. This can be interpreted as the outcome of a hypothetical game in which a malevolent player chooses a model (subject to a penalty for unreasonable models) to minimize the expected utility that the investor derives from each possible decision rule, while the investor chooses a decision rule to maximize expected utility subject to the model choices of the malevolent player. Alternatively, it can be justified intuitively as a way to find choices that will perform well within a set of models in the neighborhood of the reference model. In the words of Cagetti, Hansen, Sargent, and Williams (2002), “As a vehicle to explore the directions in which a candidate decision rule is most fragile, the decision maker considers the model that delivers the worst utility. We are not saying that he believes that the true model is the worst one, but that by planning a best response against it, he can devise a rule that is robust under a set of models surrounding his approximating [reference] model.”

A key issue is how the set of alternative models is constrained. Hansen and Sargent use relative entropy to measure the deviation of a model from the reference model. In a single-period discrete-state setting, if the reference model has state probabilities \( \pi_s \) and an alternative model has probabilities \( \pi_s' \), the relative entropy of the alternative model is defined as

\[
- \sum_{s=1}^{S} \pi_s' \log \left( \frac{\pi_s'}{\pi_s} \right) = -E \left[ \log \left( \frac{\pi'}{\pi} \right) \right] = -E \left[ \left( \frac{\pi'}{\pi} \right) \log \left( \frac{\pi'}{\pi} \right) \right],
\]

which is always positive and increases with the variability of state-probability deviations between the two models. Hansen and Sargent impose a penalty on the malevolent player that is increasing in relative entropy, thereby keeping the model chosen by the malevolent player in the neighborhood of the reference model.

To understand the implications of this framework, consider a static portfolio choice problem with a risky asset and a safe asset, where the reference model assigns a positive risk premium to the risky asset. If an investor chooses a large position in the risky asset, the malevolent agent will choose a model with a very low or even negative risk premium, implying very low expected utility. A small position in the risky asset, however, reduces the gain to the malevolent agent from distorting the risk premium (and in the extreme of a zero risky-asset position, there would be no gain). Hence, given the entropy penalty on
the malevolent agent, the solution features a small positive investment in the risky asset, just as if the investor were more risk-averse. In a dynamic setting, Maenhout (2004) and Hansen and Sargent (2008) show that this approach provides an alternative micro-foundation for Epstein-Zin (1989) preferences where the coefficient of relative risk aversion exceeds the inverse of the elasticity of intertemporal substitution. An alternative implementation of ambiguity aversion, due to Gilboa and Schmeidler (1989) and Epstein and Wang (1994), can even generate nonparticipation in risky asset markets which would require infinite risk aversion in a standard model.

The literature on ambiguity aversion blurs the distinctions between positive and normative economics, and between rational and irrational decision-making. Conservative pessimism can be treated as a positive prediction about investor behavior, but it can also be defended as a normatively justifiable (“robust”) response to model uncertainty. This contrasts with the behavioral finance literature, in which some investors are regarded as having beliefs or behaviors that do objective damage to their interests.

Related to this, Hansen and Sargent suggest that a statistical approach can be used to choose the size of the entropy penalty in their framework. The worst-case model can be chosen to be rejectable at some critical value, say 5%, given a sample of a fixed size generated by the reference model. In this way the deviation of investor beliefs from the reference model can be kept at a normatively defensible level given the data available to the investor. This, and the related issue of how conservatism varies over time in a dynamic model, are active areas of current research.

Both the literature on ambiguity and the behavioral finance literature model asset prices using assumed deviations from rational expectations. However, the literature on ambiguity assumes that investors are always more pessimistic than a rational agent would be. An old Wall Street adage describes financial markets as “the balance of fear and greed”. Hansen and Sargent’s research on robust optimal control emphasizes belief distortions driven by fear. Some behavioral finance models, such as Benartzi and Thaler’s (1995) model of myopic loss aversion, are similar in spirit, but others, such as Brunnermeier and Parker’s (2005) model of anticipation utility, or Shiller’s (2000) discussion of “irrational exuberance”, emphasize belief distortions driven by greed.

6 Practical Consequences

The 2013 Nobel laureates have had immense influence on financial practice, as well as on the academic understanding of finance. In this section I review several ways in which their work has affected, and will continue to affect, the financial industry.
6.1 Market efficiency, passive investing, and factor models

Eugene Fama’s early work on market efficiency contributed to the spread of low-cost passive investing through index funds. The first index fund was launched in 1971, shortly after the publication of Fama’s (1970) survey, and Bernstein (1993) describes the many connections between academic finance economists and innovators in the mutual fund industry. The claim that a value-weighted equity index is mean-variance efficient is one of three main arguments for passive investing. The second is the observation—which is true in any market, efficient or not—that the average return to active investing, before fees, must be exactly the same as the average return on the value-weighted index, because the profits of some active investors come from the losses of others. The third argument for passive investing, which has attracted more academic attention in recent years (Berk and Green 2002), is the claim that managers who have sufficient skill to beat an index fund will charge fees that extract all the rents from their skill, leaving nothing over for investors. French (2008) reviews this literature.

Fama later came to take a more negative view of the CAPM, finding low rewards to beta in the cross-section of stock returns (Fama and French 1992). He responded to this by introducing the famous Fama-French three-factor model (Fama and French 1993), which adds two new factors to the market factor, long-short portfolios that go long value stocks with high book-market ratios and short growth stocks with low book-market ratios (HML) and go long small stocks and short large or “big” stocks (SMB).

The idea of a multi-factor model cannot be credited to Fama and French. Merton (1973) and Ross (1976) provide theoretical foundations for multi-factor asset pricing, and a large empirical literature on multi-factor models is summarized by Campbell, Lo, and MacKinlay (1997, Chapter 6). Fama and French’s contribution is to create factors by sorting on stock characteristics that correlate with expected returns. To the extent that such sorts capture genuine drivers of expected returns and not spurious small-sample phenomena, then Ross’ arguments imply that this procedure is a legitimate alternative to earlier approaches that directly analyze the structure of the variance-covariance matrix of stock returns. Empirically, the Fama-French approach has been an effective way to simplify and unify the vast literature on the cross-section of stock returns (Cochrane 2011).

Academics following Fama and French have explored other data samples (Davis, Fama, and French 2000), have suggested additional factors, such as the momentum factor of Carhart (1997), and have asked why risk factors such as HML and SMB should have nonzero prices. Why should investors care about exposure to the common movements of value stocks or small stocks? Fama and French (1996) presents informal arguments that value stocks may be correlated with human capital, an omitted component of wealth in traditional implementations of the CAPM. However Fama has left more formal empirical analysis of multifactor risk pricing to others, and both mainstream and behavioral finance researchers have taken up this challenge as discussed above. There are multiple rational models that explain the value premium, but the reward for momentum is much more challenging to explain in a
rational framework and has inspired behavioral models.

In the financial industry, Fama and French's portfolios are influential in a different way. They are useful in risk management because stocks with these characteristics tend to move together. They are also useful for constructing low-cost investment strategies, as the concept of an index fund has been extended from its original application to value-weighted market indexes to include style-based value and size indexes. Institutional investors have become increasingly wary of paying high active management fees for exposures to Fama-French style tilts, and asset managers have responded by offering so-called “smart beta” products that offer exposures to historically profitable styles at low cost.

6.2 Quantitative investing

The term “quant” means different things in different parts of the financial industry. To investors in fixed-income securities and their derivatives, quants are experts in pricing and dynamically hedging nonlinear exposures to systematic risks. They are often former mathematicians or physicists with strong skills in numerical solution methods for partial differential equations. Works such as Derman (2007) and Blyth (2013) use the word in this sense. To investors in equities, however, quants are followers of Fama, Hansen, and Shiller who combine the insights of behavioral finance with the tools of modern financial econometrics, all the while mindful of the basic insight of market efficiency that easily identifiable profit opportunities tend to disappear.

Three firms that practice quantitative investing in this latter sense are Dimensional Fund Advisors, founded by Fama’s students Rex Sinquefield and David Booth (who later endowed and renamed the Chicago Graduate School of Business as the Booth School of Business); AQR, co-founded by Fama’s student Clifford Asness; and Arrowstreet Capital, of which I am a co-founder. Dimensional and AQR are described in Harvard Business School cases by Cohen (2002) and Bergstresser, Cohen, Cohen, and Malloy (2010). People associated with all three firms continue to publish actively in empirical asset pricing, and were among those gathered in Stockholm in December 2013. It is a fascinating irony that when academics believed most fervently in simple efficient-markets models with constant discount rates, during the 1960s and 1970s, they made no effort to build businesses that would help to enforce efficient pricing through arbitrage trading. Only later, when empirical evidence accumulated that complicated the academic worldview, did academics become involved in this activity and thereby contribute directly to the elimination of at least some capital markets anomalies.
6.3 Financial innovation

Robert Shiller is well known for his view that speculative capital markets are subject to long waves of irrational optimism and pessimism that cause prices to deviate substantially from the levels that would prevail if all investors were rational utility-maximizers. On two occasions he has publicly argued that stock prices (Campbell and Shiller 1998, Shiller 2000) or house prices (Shiller 2005) were overvalued, and has made real-time forecasts of price declines. He has also written about the macroeconomic effects of such “animal spirits” (Akerlof and Shiller 2010).

However this view does not lead Shiller to be hostile to modern financial capitalism, or to call for strict regulation of existing capital markets. Instead, in a series of books written for a general audience, Shiller has argued that economic progress requires the development of new financial markets that are better designed to allow investors to share risks and to focus their thinking on fundamental uncertainty.

Shiller’s work in this area ranges from analyses of inflation-indexed bonds, which are financial instruments with a long history although they were not issued by the US Treasury until 1997, through housing markets where the challenge is to develop price indexes suitable for the development of derivative securities, to more radical proposals for new financial instruments that would allow households to share income risks across countries, sectors, and occupations.

Two papers on inflation-indexed bonds (Campbell and Shiller 1996, Campbell, Shiller, and Viceira 2009) argue for the importance of these instruments. In a world of incomplete markets, there are many missing markets that can be created, but inflation-indexed bonds are long-term safe assets that should dominate the portfolios of conservative long-term investors (Campbell and Viceira 2001). Campbell and Shiller (1996) emphasize that inflation-indexation of government debt should encourage inflation-indexation of financial contracts, reducing arbitrary transfers across agents that result from inflation surprises when contracts are nominal, and should encourage imperfectly rational investors to think about the distinction between nominal and real interest rates, thereby reducing the mispricing that can result from inflation illusion (Modigliani and Cohn 1979). In fact, while Treasury inflation-protected securities (TIPS) are now widely accepted, there has been little inflation-indexation elsewhere in the economy, possibly because inflation expectations have been relatively stable since the early 2000s, or because people find it cognitively challenging to think in real rather than nominal terms.

Shiller’s work on housing began in the 1980s in collaboration with Karl Case. Case and Shiller (1989, 2003) showed how to construct repeat-sales indexes that control for variation in the unobserved quality of homes sold at different times, and that can be constructed at a monthly frequency even though individual houses sell relatively infrequently. Case and Shiller, together with Allan Weiss, founded a company Case Shiller Weiss, Inc. in 1991 to commercialize this idea. During the 2000s, Case-Shiller house price indexes were widely
publicized and played an important role in discussions of house price increases before the onset of the global financial crisis. Shiller (2005) used them as the basis for his public statements that houses were overvalued in the early 2000s. Working with a new company, MacroMarkets LLC, Case and Shiller supported the launch in 2006 of housing futures markets based on the Case-Shiller indexes, and tried to create securities tied to these indexes (Greenwood and Viceira 2010). Most recently, Shiller has argued for “continuous-workout” mortgages whose principal balance would vary along with Case-Shiller house price indexes (Shiller 2008).

Shiller’s most visionary work on new financial instruments proposes the creation of derivatives markets that can be used to hedge important macroeconomic risks, along the lines of CPI futures that were briefly traded in the mid-1980s at the instigation of Paul Samuelson’s student and commodity trader Helmut Weymar. At the theoretical level, Athanasoulis and Shiller (2000) solve the market design problem of a social planner who can create only a limited number of financial contracts, too few to permit complete risk-sharing among agents with heterogeneous income risks. They show that the optimal contracts look like swaps, with positive loadings on some people’s incomes and negative loadings on others. In fact, these loadings are orthogonal to the loadings of the market portfolio that pays off the sum of all incomes and thus has equal loadings on each person’s income. Athanasoulis and Shiller (2001) is an illustrative empirical analysis of cross-country risk-sharing, which shows how beneficial it would be to trade US-Japanese or Japanese-European income swaps. Shiller has promoted these ideas in two books written for a general audience (2004, 2012).

7 Conclusion

The 2013 Nobel laureates are three economists with their own unique perspectives and research styles. They disagree on some important questions, such as the relevance of irrational expectations for asset prices. Fama doubts that we need to consider any form of irrationality, while Shiller emphasizes irrational exuberance and Hansen models a form of conservative pessimism. The laureates also write distinctively. Fama’s papers are justly famous for their summaries of the key properties of complex financial datasets, while his interpretations of the data are largely verbal. Hansen is a master of contemporary formal modeling. Shiller has written many rigorous research papers, but he often cites writings in other disciplines and seeks to persuade a broad audience of the merit of his ideas.

Despite this diversity of views and approaches, I have argued in this paper that the 2013 economics award possesses a deep intellectual coherence. Empirical research in asset pricing today is conducted within a common intellectual framework, due in large part to the methodological contributions of the laureates. Important themes in the literature can best be understood by considering the laureates in pairs. I have discussed the joint hypothesis problem and opportunity in testing models of market efficiency (Fama and Hansen); the
predictability of asset returns in the short and long run (Fama and Shiller); and the modeling of optimistic and pessimistic investor beliefs (Shiller and Hansen).

Empirical economists working in asset pricing, including the laureates and their students, have documented a rich variety of facts about asset prices. For example, news events typically move asset prices in a manner that scales appropriately with the fundamental impacts of the events, allowing asset prices to be used to indirectly measure such fundamental impacts. Asset prices sometimes drift in the aftermath of events, most famously corporate earnings announcements, but these drifts typically weaken over time as arbitrageurs exploit them. Aggregate fluctuations in asset prices appear to reflect variation in discount rates, and specifically risk premia, so that valuation ratios can be used to predict returns. Time-varying risk premia do not move in proportion with return volatility, and they tend to be countercyclical, rising when the economy deteriorates. In the cross-section of stock returns, extremely modest predictability in the returns of individual stocks can be amplified by sorting stocks with similar characteristics into portfolios. Portfolios of value stocks and momentum stocks, to take the two most famous examples, have high average returns. High returns to these characteristics are found in other asset classes as well, and rewards for asset-class risk exposures are typically higher when the risk is taken by leveraging an index than when it is taken by buying individual assets with high betas to the index. These facts are reviewed and summarized in papers such as Asness, Moskowitz, and Pedersen (2013) and Cochrane (2011), academic books such as Campbell, Lo, and MacKinlay (1997) and Cochrane (2005), and trade books such as Ilmanen (2011) and Siegel (1994). Taken together, they make asset pricing one of the most successful empirical fields in economics.

Research in asset pricing has also had an extraordinary impact on the practice of finance. From the growth of index funds in the 1970s and the use of efficient market theory in securities litigation, through the more recent development of quantitative asset management, to the introduction of house price indexes and associated derivatives and securities, the financial industry has drawn upon the work of the 2013 laureates. Financial innovators continue to be inspired by the work of these great economists.
References


