Two Pillars of Asset Pricing

Prize Lecture, December 8, 2013

by Eugene F. Fama
Booth School, University of Chicago, Chicago, IL, USA.

The Nobel Foundation asks that the Prize lecture cover the work for which the Prize is awarded. The announcement of this year’s Prize cites empirical work in asset pricing. I interpret this to include work on efficient capital markets and work on developing and testing asset pricing models—the two pillars, or perhaps more descriptive, the Siamese twins of asset pricing. I start with efficient markets and then move on to asset pricing models.

EFFICIENT CAPITAL MARKETS

A. Early Work

The year 1962 was a propitious time for Ph.D. research at the University of Chicago. Computers were coming into their own, liberating econometricians from their mechanical calculators. It became possible to process large amounts of data quickly, at least by previous standards. Stock prices are among the most accessible data, and there was burgeoning interest in studying the behavior of stock returns, centered at the University of Chicago (Merton Miller, Harry Roberts, Lester Telser, and Benoit Mandelbrot as a frequent visitor) and MIT (Sidney Alexander, Paul Cootner, Franco Modigliani, and Paul Samuelson). Modigliani often visited Chicago to work with his longtime coauthor Merton Miller, so there was frequent exchange of ideas between the two schools.

It was clear from the beginning that the central question is whether asset prices reflect all available information—what I labeled the efficient markets hypothesis (Fama 1965b). The difficulty is making the hypothesis testable. We can’t
test whether the market does what it is supposed to do unless we specify what
it is supposed to do. In other words, we need an asset pricing model, a model
that specifies the characteristics of rational expected asset returns in a market
equilibrium. Tests of efficiency basically test whether the properties of expected
returns implied by the assumed model of market equilibrium are observed in
actual returns. If the tests reject, we don't know whether the problem is an inef-
fficient market or a bad model of market equilibrium. This is the joint hypothesis

A bit of notation makes the point precise. Suppose time is discreet, and \( P_{t+1} \)
is the vector of payoffs at time \( t + 1 \) (prices plus dividends and interest pay-
ments) on the assets available at \( t \). Suppose \( f(P_{t+1} | \Theta_{tm}) \) is the joint distribution
of asset payoffs at \( t + 1 \) implied by the time \( t \) information set \( \Theta_{tm} \) used in the
market to set \( P_t \), the vector of equilibrium prices for assets at time \( t \). Finally, sup-
pose \( f(P_{t+1} | \Theta_t) \) is the distribution of payoffs implied by all information available
at \( t \), \( \Theta_t \); or more pertinently, \( f(P_{t+1} | \Theta_t) \) is the distribution from which prices at
\( t + 1 \) will be drawn. The market efficiency hypothesis that prices at \( t \) reflect all
available information is,

\[
f(P_{t+1} | \Theta_{tm}) = f(P_{t+1} | \Theta_t).
\]

The market efficiency condition is more typically stated in terms of expected
returns. If \( E(R_{t+1} | \Theta_{tm}) \) is the vector of expected returns implied by \( f(P_{t+1} | \Theta_{tm}) \)
and the equilibrium prices \( P_t \), and \( E(R_{t+1} | \Theta_t) \) is the expected return vector im-
plied by time \( t \) prices and \( f(P_{t+1} | \Theta_t) \), the market efficiency condition is,

\[
E(R_{t+1} | \Theta_{tm}) = E(R_{t+1} | \Theta_t),
\]

The prices observed at \( t + 1 \) are drawn from \( f(P_{t+1} | \Theta_t) \), so in this sense
\( f(P_{t+1} | \Theta_t) \) and \( E(R_{t+1} | \Theta_t) \) are observable, but we do not observe \( f(P_{t+1} | \Theta_{tm}) \)
and \( E(R_{t+1} | \Theta_{tm}) \). As a result, the market efficiency conditions (1) and (2) are
not testable. To have testable propositions, we must specify how equilibrium
prices at \( t \) relate to the characteristics of \( f(P_{t+1} | \Theta_{tm}) \). In other words, we need
a model of market equilibrium—an asset pricing model, no matter how primi-
tive—that specifies the characteristics of rational equilibrium expected returns,
\( E(R_{t+1} | \Theta_{tm}) \).

For example, in many early tests, market efficiency is assumed to imply that
returns are unpredictable based on past information. The implicit model of
market equilibrium is that equilibrium expected returns are constant,

\[
E(R_{t+1} | \Theta_{tm}) = E(R).
\]
If the market is efficient so that (2) holds, then

\[ E(R_{t+1} | \Theta_t) = E(R). \]

The testable implication of (4) is that a regression of \( R_{t+1} \) on variables from \( \Theta_t \), which are known at time \( t \), should produce slopes that are indistinguishable from zero. If the test fails, we don’t know whether the problem is a bad model of market equilibrium (equation (3) is the culprit) or an inefficient market that overlooks information in setting prices (equations (1) and (2) do not hold). This is the joint hypothesis problem.

The joint hypothesis problem is perhaps obvious in hindsight, and one can argue that it is implicit in Bachelier (1900), Muth (1961), Samuelson (1965), and Mandelbrot (1966). But its importance in work on market efficiency was not recognized before Fama (1970), which brought it to the forefront.

For example, many early papers focus on autocorrelations and it was common to propose that market efficiency implies that the autocorrelations of returns are indistinguishable from zero. The implicit model of market equilibrium, never acknowledged in the tests, is (3), that is, the market is trying to price stocks so that their expected returns are constant through time.

A clean statement of the joint hypothesis problem, close to that given above, is in Chapter 5 of Fama (1976b). Everybody in finance claims to have read this book, but given its sales, they must be sharing the same copy.

Market efficiency is always tested jointly with a model of market equilibrium, but the converse is also true. Common asset pricing models, like the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), Merton’s (1973a) intertemporal CAPM (the ICAPM), and the consumption CAPM of Lucas (1978) and Breeden (1979), implicitly or explicitly assume that all information is costlessly available to all market participants who use it correctly in their portfolio decisions—a strong form of market efficiency. Thus, tests of these asset pricing models jointly test market efficiency.

### B. Event Studies

In the initial empirical work on market efficiency, the tests centered on predicting returns using past returns. Fama, Fisher, Jensen, and Roll (FFJR 1969) extend the tests to the adjustment of stock prices to announcements of corporate events. In FFJR the event is stock splits, but the long-term impact of the paper traces to the empirical approach it uses to aggregate the information about price adjustment in a large sample of events.
Like other corporate events, the sample of splits is spread over a long period (1926–1960). To abstract from general market effects that can obscure a stock's response to a split, we use a simple “market model” time series regression,

\[ R_{it} = a_i + b_i R_{Mt} + e_{it}. \]

In this regression, \( R_{it} \) is the return on stock \( i \) for month \( t \), \( R_{Mt} \) is the market return, and the residual \( e_{it} \) is the part of the security's return that is not a response to the market return. The month \( t \) response of the return to a split is thus embedded in \( e_{it} \). To aggregate the responses across the stocks that experience a split, we use event time rather than calendar time. Specifically, \( t = 0 \) is the month when information about a split becomes available, \( t = -1 \) is the previous month, \( t = 1 \) is the following month, etc. Thus, period 0 is a different calendar month for each split. To measure the average response of returns in the months preceding and following a split, we average the residuals for the stocks in the sample for each of the 30 months preceding and following the split. To measure the cumulative response, we sequentially sum the average residuals.

The results of the split paper are striking. The cumulative average residual (Figure 1) rises in the months preceding a split. Thus, companies tend to split their stocks after good times that produce large increases in their stock prices. Once the split becomes known, however, there is no further movement in the cumulative average residual, despite the fact that about 75% of the companies that split their stocks continue to experience good times (witnessed by subsequent dividend growth rates larger than those of the market as a whole). In other words, on average, all the implications of a split for the future performance of a company are incorporated in stock prices in the months leading up to the split, with no further reaction thereafter—exactly the prediction of market efficiency.

The split paper spawned an event study industry. To this day, finance and accounting journals contain many studies of the response of stock prices to different corporate events, for example, earnings announcements, merger announcements, security issues, etc. Almost all use the simple methodology of the split paper. Like the split study, other early event studies generally confirm that the adjustment of stock prices to events is quick and complete.

Early event studies concentrate on short periods, typically days, around an event. Over short periods the assumed model for equilibrium expected returns is relatively unimportant because the change in the price of the stock in response to the event is typically much larger than short horizon expected returns. In other words, the joint hypothesis problem is relatively unimportant. More recently, researchers in behavioral finance became interested in studying price
responses for several years after an event. Over such long periods, expected returns are larger relative to the price effect of the event, and the joint hypothesis problem becomes important.

For example, the implicit model of market equilibrium in the split study is that the regression intercept and slope, \( a_i \) and \( b_i \), in the market model regression (5) are constant through time. It is now well-known that \( a_i \) and \( b_i \) change through time. This can produce drift in long-term cumulative average regression residuals that looks like market inefficiency but is just a bad model for expected returns. These issues are discussed in Fama (1998).

### C. Predictive Regressions

The early work on market efficiency focuses on stock returns. In Fama (1975), I turn to bonds to study Irving Fisher’s (market efficiency) hypothesis that \( i_{t+1} \), the time \( t \) interest rate on a short-term bond that matures at \( t + 1 \), should contain the equilibrium expected real return, \( E(r_{t+1}) \), plus the best possible forecast of the inflation rate, \( E(\pi_{t+1}) \),

\[
i_{t+1} = E(r_{t+1}) + E(\pi_{t+1}).
\]
The topic is not new, but my approach is novel. Earlier work uses regressions of the interest rate on lagged inflation rates,

\[ i_{t+1} = a + b_1\pi_t + b_2\pi_{t-1} + \ldots + \varepsilon_{t+1}. \]

The idea is that the expected inflation rate (along with the expected real return) determines the interest rate, so the interest rate should be the dependent variable and the expected inflation rate should be the independent variable. Past inflation is a noisy measure of expected inflation, so equation (7) suffers from an errors-in-variables problem. More important, in an efficient market the expected inflation rate built into the interest rate surely uses more information than past inflation rates.

The insight in Fama (1975), applied by me and others in subsequent papers, is that a regression estimates the conditional expected value of the left-hand-side variable as a function of the right-hand-side variables. Thus, to extract the forecast of inflation in the interest rate (the expected value of inflation priced into the interest rate), one regresses the inflation rate for period \( t + 1 \) on the interest rate for \( t + 1 \) set at the beginning of the period,

\[ \pi_{t+1} = a + b_i_{t+1} + \varepsilon_{t+1}. \]

The expected inflation rate estimated in this way captures all the information used to set the interest rate. In hindsight, this is the obvious way to run the forecasting regression, but it was not obvious at the time.

Reversing the regression eliminates one measurement error problem, but it can introduce another, caused by variation through time in the expected real return built into the interest rate. The model of market equilibrium in Fama (1975) is that the expected real return is constant, \( E(r_{t+1}) = r \). Near zero autocorrelations of real returns suggest that this proposition is a reasonable approximation, at least for the 1953–1971 period examined. Thus, at least for this period, the interest rate \( i_{t+1} \) is a direct proxy for the expected inflation rate—it is the expected inflation rate plus a constant.

The slopes in the estimates of (8) for the one-month, three-month, and six-month U.S. Treasury Bill rates and inflation rates of 1953–1971 are quite close to 1.0, and the autocorrelations of the residuals are close to zero. Thus, the bottom line from Fama (1975) is that interest rates on one-month, three-month, and six-month Treasury Bills seem to contain rational forecasts of inflation one, three, and six months ahead.

Fisher’s hypothesis that expected asset returns should include compensation for expected inflation applies to all assets. Fama and Schwert (1977) test it on
longer term bonds, real estate, and stock returns. The proposed model of market equilibrium has two parts. First, as in Fama (1975), the equilibrium expected real returns on bills are assumed to be constant through time, so the bill rate can again be used as the proxy for the expected inflation rate. Second, any variation in equilibrium expected real returns on other assets is assumed to be uncorrelated with expected inflation. With this model of market equilibrium, we can test Fisher’s hypothesis with regressions of the nominal return on an asset, \( R_{t+1} \), on the bill rate set at the beginning of period \( t + 1 \),

\[
R_{t+1} = a + b i_{t+1} + \epsilon_{t+1}.
\]  

The tests say that monthly, quarterly, and semi-annual nominal returns on longer-term bonds and real estate compensate for monthly, quarterly, and semi-annual expected inflation: that is, for these assets the slopes in the estimates of (9) are again near 1.0. Thus, we cannot reject the market efficiency proposition that bond and real estate prices incorporate the best possible forecasts of inflation and the model of market equilibrium in which expected real returns vary independently of expected inflation.

The relation between common stock returns and expected inflation, however, is perverse. The slopes in the estimates (9) for stocks are negative; expected stock returns are higher when expected inflation (proxied by the bill rate) is low and vice versa. Thus for stocks we face the joint hypothesis problem. Do the tests fail because of poor inflation forecasts (market inefficiency) or because equilibrium expected real stock returns are in fact negatively related to expected inflation (so we chose a bad model of market equilibrium)?

The simple idea about forecasting regressions in Fama (1975)—that the regression of a return on predetermined variables produces estimates of the variation in the expected value of the return conditional on the forecasting variables—has served me well. I have used it in a sequence of papers to address an old issue in the term structure literature, specifically, how well do the forward interest rates that can be extracted from prices of longer-term discount bonds forecast future one-period (spot) interest rates (Fama 1976a,c, 1984b,c, 1986, 1990a, 2005, and Fama and Bliss 1987).

To see the common insight in these term structure papers, define the term (or maturity) premium in the one-period return on a discount bond with \( T \) periods to maturity at time \( t \) as the difference between the return, \( RT_{t+1} \), and the one-period “spot” interest rate observed at time \( t \), \( S_{t+1} \). Skipping the tedious details, it is easy to show that the time \( t \) forward rate for period \( t + T \), \( F_{t+T} \), contains the expected term premium, \( E(RT_{t+1}) - S_{t+1} \), as well as a forecast of the spot
rate for \( t + T, E(S_{t+T}) \). As a result, there is a pair of complementary regressions that use the difference between the forward rate and the current spot rate to forecast the term premium and the future change in the spot rate,

\[
RT_{t+1} - S_{t+1} = a_1 + b_1(F_{t+1} - S_{t+1}) + e_{1,t+1}
\]

\[
S_{t+T} - S_{t+1} = a_2 + b_2(F_{t+1} - S_{t+1}) + e_{2,t+1}
\]

The conclusion from this work is that the information in forward rates is primarily about expected term premiums rather than future spot rates; that is, the slope in (10) is near 1.0, and the slope in (11) is near 0.0. There is, however, some longer-term predictability of spot rates due to mean reversion of the spot rate (Fama and Bliss 1987), though not necessarily to a constant mean (Fama 2005).

In Fama (1984a), I apply the complementary regression approach to study forward foreign exchange rates as predictors of future spot rates. Again, the information in forward exchange rates seems to be about risk premiums, and there is little or no information about future spot exchange rates. The exchange rate literature has puzzled over this result for 30 years. Using the complementary regression approach, Fama and French (1987) find that futures prices for a wide range of commodities do show power to forecast spot prices—the exception to the general rule.

**D. Time Varying Expected Stock Returns**

As noted above, early work on market efficiency generally assumes that equilibrium expected stock returns are constant through time. This is unlikely to be true. The expected return on a stock contains compensation for bearing the risk of the return. Both the risk and the willingness of investors to bear the risk are likely to change over time, leading to a time-varying expected return. The trick is to find predetermined variables that can be used to track expected returns in forecasting regressions.

Fama and Schwert (1977) document variation in monthly, quarterly, and semi-annual expected stock returns using predetermined monthly, quarterly, and semi-annual Treasury bill rates. In later work, the popular forecasting variable on the right hand side of the regression is the dividend yield, the ratio of trailing annual dividends to the stock price at the beginning of the forecast period. The motivation, which I attribute to Ball (1978), is that a stock's price is the present value of the stream of expected future dividends, where the discount
rate is (approximately) the expected stock return. Thus, a high stock price relative to dividends likely signals a lower expected return, and vice versa. The word “likely” is needed because price also depends on expected future dividends, which means the dividend yield is a noisy proxy for the expected stock return, a problem emphasized by Campbell and Shiller (1988) and others. Cochrane (2010) gives an elegant explanation of the problem in terms of complementary regressions that use the dividend yield to forecast long-term average stock returns and long-term dividend growth.

To my knowledge, the first papers that use dividend yields to track expected stock returns are Rozeff (1984) and Shiller (1984). Fama and French (1987) add an interesting wrinkle to the evidence. We find that the explanatory power of the regression, measured by the regression $R^2$, increases as the horizon for the return is extended in steps from a month to four years. This result may seem surprising, but it is just a consequence of the fact that dividend yields are persistent (highly autocorrelated).

For example, with persistent dividend yields, the slope in the regression of the quarterly stock return on the beginning of quarter yield will be about three times the slope in the regression of the monthly return on the beginning of month yield. Thus, the variance of the expected return estimate in the three-month regression is about nine times the variance in the one-month regression. But the variance of the residual in the three-month regression (the unexpected part of the return) is only about three times the variance of the residual in the one-month regression. As a result, $R^2$ is higher in the three-month regression.

Higher $R^2$ for longer return horizons due to the persistence of the dividend yield implies that the variance of the predictable part of returns rises faster than the variance of the unpredictable part, so in this sense longer horizon returns are more predictable. But unpredictable variation in returns also rises with the return horizon, that is, the variance of forecast errors is larger in longer-term returns, so in this more important sense, longer horizon returns are less predictable.

Efficient market types (like me) judge that predictable variation in expected returns on stocks and bonds is rational, the result of variation in risk or willingness to bear risk. In contrast, behaviorists argue that much of the predictability is due to irrational swings of prices away from fundamental values.

Fama and French (1989) address this issue. They find that the well-known variation in expected bond returns tracked by two term structure variables, (i) the default spread (the difference between the yields on long-term bonds of high and low credit risk) and (ii) the term spread (the difference between long-term and short-term yields on high grade bonds) is shared with stock returns. Likewise, dividend yields predict bond returns as well as stock returns. Moreover,
default spreads and dividend yields are related to long-term business conditions, and term spreads are strongly related to short-term business cycles. The general result is that expected returns are high when business conditions are poor and low when they are strong.

The evidence that the variation in expected returns is common to stocks and bonds and related to business conditions leads Fama and French (1989) to conclude that the resulting predictability of stock and bond returns is rational. Behaviorists can disagree. Animal spirits can roam across markets in a way that is related to business conditions. No available empirical evidence resolves this issue in a way that convinces both sides.

Shiller (1981) finds that the volatility of stock prices is much higher than can be explained by the uncertain evolution of expected future dividends. This result implies that much of the volatility of stock prices comes from time-varying expected returns. The market efficiency issue is whether the variation in expected returns necessary to explain Shiller’s results is beyond explanation by a model for rational expected returns. It is certainly possible to develop models for expected returns that produce this conclusion in empirical tests. But then we face the joint hypothesis problem. Do the tests fail because the market is inefficient or because we have the wrong model for rational expected returns? This and other market efficiency issues are discussed in detail in Fama (1991).

E. “Bubbles”

There is one remaining result in the literature on return predictability that warrants mention. The available evidence says that stock returns are somewhat predictable from dividend yields and interest rates, but there is no statistically reliable evidence that expected stock returns are sometimes negative. Fama and French (1987) find that predictions from dividend yields of negative returns for market portfolios of U.S stocks are never more than two standard errors below zero. Fama and Schwert (1977) find no evidence of reliable predictions of negative market returns when the forecast variable is the short-term bill rate.

These results are important. The stock market run-up to 2007 and subsequent decline is often called a “bubble.” Indeed, the word “bubble,” applied to many markets, is now common among academics and practitioners. A common policy prescription is that the Fed and other regulators should lean against asset market bubbles to preempt the negative effects of bursting bubbles on economic activity.

Such policy statements seem to define a “bubble” as an irrational strong price increase that implies a predictable strong decline. This also seems to be
the definition implicit in most recent claims about “bubbles.” But the available research provides no reliable evidence that stock market price declines are ever predictable. Thus, at least as the literature now stands, confident statements about “bubbles” and what should be done about them are based on beliefs, not reliable evidence.

“Reliable” is important in this discussion. After an event, attention tends to focus on people who predicted it. The *ex post* selection bias is obvious. To infer reliability for a particular forecaster, we need to evaluate his or her track record of forecasts of different events, and for a particular event, we must evaluate the initial predictions of all forecasters we might have chosen *ex ante*. More important, for the purposes of science and policy, we are interested in the performance of forecasting models, not the unreproducible predictions of specific individuals.

The absence of evidence that stock market price declines are predictable seems sufficient to caution that “bubble” is a treacherous term, but there is more. Figure 2 shows the December 1925 to September 2013 path of the natural log of U.S. stock market wealth, including reinvested dividends, constructed using the value-weight market portfolio of NYSE, AMEX, and NASDAQ stocks from the Center for Research in Security Prices (CRSP) of the University of Chicago. The recessions identified by the NBER are shown as shaded areas on the graph.

![Figure 2](image-url)

**Figure 2.** Log of cumulative value of the CRSP market index, including dividends. Shaded areas are U.S. recessions identified by the National Bureau of Economic Research (NBER).
In percent terms, and noting that these are end-of-month data, the largest five price declines in Figure 2 are (1) August 1929 to June 1932, (2) October 2007 to February 2009, (3) February 1937 to March 1938, (4) August 2000 to September 2002, and (5) August 1972 to December 1974. All these price declines are preceded by strong price increases, so these are prime “bubble” candidates.

These five periods are associated with recessions, and except for August 2000 to September 2002, the magnitude of the price decline seems to reflect the severity of the recession. The peak of the market in 1929 is the business cycle peak, but for the other four episodes, the market peak precedes the business cycle peak. Except for August 2000 to September 2002, the market low precedes the end of the recession. This pattern in stock prices also tends to occur around less severe recessions.

It thus seems that large swings in stock prices are responses to large swings in real activity, with stock prices forecasting real activity—a phenomenon studied in detail in Fama (1981, 1990b). All this is consistent with an efficient market in which the term “bubble,” at least as commonly used, has no content.

One might assert from Figure 2 that major stock market swings cause recessions and market upturns bring them to an end. (One can also assert that the weatherman causes the weather—a quip stolen from John Cochrane.) At a minimum, however, (i) the absence of evidence that price declines are ever predictable, and (ii) the evidence that the prime “bubble” candidates seem to be associated with rather impressive market forecasts of real activity are sufficient to caution against use of the “bubble” word without more careful definition and empirical validation.

Common “bubble” rhetoric says that the declines in prices that terminate “bubbles” are market corrections of irrational price increases. Figure 2 shows, however, that major stock price declines are followed rather quickly by price increases that wipe out, in whole or in large part, the preceding price decline. “Bubble” stories thus face a legitimate question: which leg of a “bubble” is irrational, the up or the down? Do we see irrational optimism in the price increase corrected in the subsequent decline? Or do we see irrational pessimism in the price decline, quickly reversed? Or both? Or perhaps neither?

Finally, it is difficult to evaluate expert forecasts of “bubbles” in asset prices since we tend to hear only “success” stories identified after the fact, and for a particular “bubble,” we rarely know the all-important date of an expert’s first forecast that prices are irrationally high. For a bit of fun, however, we can examine two commonly cited “success” stories.

On the website for his book, Irrational Exuberance, Shiller says that at a December 3, 1996 lunch, he warned Fed chairman Alan Greenspan that the level of
stock prices was irrationally high. Greenspan’s famous “Irrational Exuberance” speech followed two days later. How good was Shiller’s forecast? On December 3, 1996 the CRSP index of U.S. stock market wealth stood at 1518. It more than doubled to 3191 on September 1, 2000, and then fell. This is the basis for the inference that the original bubble prediction was correct. At its low on March 11, 2003, however, the index, at 1739, was about 15% above 1518, its value on the initial “bubble” forecast date. These index numbers include reinvested dividends, which seem relevant for investor evaluations of “bubble” forecasts. If one ignores dividends and focuses on prices alone, the CRSP price index on March 11, 2003 was also above its December 3, 1996 value (648 versus 618). In short, there is not much evidence that prices were irrationally high at the time of the 1996 forecast, unless they have been irrationally high ever since.

The second “success” story is the forecast in the mid-2000s that real estate prices were irrationally high. Many academics and practitioners made the same forecast, but an easy one to date is Case and Shiller (2003), which was probably written in late 2002 or early 2003. To give their prediction a good shot, I choose July 2003 as the date of the first forecast of a real estate “bubble.” The S&P/Case Shiller 20-City Home Price Index is 142.99 in July 2003, its peak is 206.52 in July 2006, and its subsequent low is 134.07 in March 2012. Thus, the price decline from what I take to be the first forecast date is only 6.7%. The value to homeowners from housing services during the almost nine years from July 2003 to March 2012 surely exceeds 6.7% of July 2003 home values. Moreover, on the last sample date, October 2013, the real estate index, at 165.91, is 16% above its value on the initial “bubble” forecast date. Again, there is not much evidence that prices were irrationally high at the time of the initial forecast.

I single out Shiller and Case and Shiller (2003) only because their initial forecasts of these two “bubbles” are relatively easy to date. Many academics, including (alas) some of my colleagues, made the same “bubble” claims at similar times, or earlier.

**F. Behavioral Finance**

I conclude this section on market efficiency with a complaint voiced in my review of behavioral finance 15 years ago (Fama 1998). The behavioral finance literature is largely an attack on market efficiency. The best of the behaviorists (like my colleague Richard Thaler) base their attacks and their readings of the empirical record on findings about human behavior in psychology. Many others don’t bother. They scour databases of asset returns for “anomalies” (a statistically treacherous procedure), and declare victory for behavioral finance when they
find a candidate. Most important, the behavioral literature has not put forth a full blown model for prices and returns that can be tested and potentially rejected—the acid test for any model proposed as a replacement for another model.

**Asset Pricing Models**

This year’s Nobel award cites empirical research in asset pricing. Tests of market efficiency are one branch of this research. The other is tests of asset pricing models, that is, models that specify the nature of asset risks and the relation between expected return and risk. Much of my work is concerned with developing and testing asset pricing models, the flip side of the joint hypothesis problem.

**A. Fama and MacBeth (1973)**

The first formal model of market equilibrium is the CAPM (capital asset pricing model) of Sharpe (1964) and Lintner (1965). In this model market β, the slope in the regression of an asset’s return on the market return, is the only relevant measure of an asset’s risk, and the cross-section of expected asset returns depends only on the cross-section of asset βs.

In the early literature, the common approach to test this prediction was cross-section regressions of average security or portfolio returns on estimates of their βs and other variables. Black, Jensen and Scholes (1972) criticize this approach because it produces estimates of the slope for β (the premium in expected returns per unit of β) that seem too precise, given the high volatility of market returns. They rightly suspect that the problem is cross-correlation of the residuals in the regression, which leads to underestimated standard errors. They propose a complicated portfolio approach to solve this problem.

Fama and MacBeth (1973) provide a simple solution to the cross-correlation problem. Instead of regressing average asset returns on βs and other variables, one does the regression period-by-period, where the period is usually a month. The slopes in the regression are monthly portfolio returns whose average values can be used to test the CAPM predictions that the expected β premium is positive and other variables add nothing to the explanation of the cross-section of expected returns. (This is best explained in chapter 9 of Fama 1976b).

An example is helpful. Fama and French (1992) estimate month-by-month regressions of the cross-section of individual stock returns for month t, \( R_{it} \), on estimates \( b_i \) of their βs, their (logged) market capitalizations at the beginning of month t, \( MC_{it-1} \), and their book-to-market equity ratios, \( BM_{it-1} \).

\[
R_{it} = a_i + a_1 b_i + a_2 MC_{it-1} + a_3 BM_{it-1} + e_{it}
\]
In the CAPM the cross-section of expected returns is completely described by the cross-section of $\beta$s, so $MC_{i,t-1}$ and $BM_{i,t-1}$ should add nothing to the explanation of expected returns. The average values of the slopes $a_{2t}$ and $a_{3t}$ for $MC_{i,t-1}$ and $BM_{i,t-1}$ test this prediction, and the average value of the slope $a_{1t}$ for $b_i$ tests the CAPM prediction that the premium for $\beta$ is positive.

The key to the test is the simple insight that the month-by-month variation in the regression slopes (which is, in effect, repeated sampling of the slopes) captures all the effects of the cross-correlation of the regression residuals (and of multicollinearity of the explanatory variables). The time-series standard errors used to calculate t-statistics for the average slopes thus capture the effects of residual covariance without requiring an estimate of the residual covariance matrix. And inferences lean on the relatively robust statistical properties of $t$ tests for sample means.

The Fama-MacBeth approach is standard in tests of asset pricing models that use cross-section regressions, but its benefits carry over to panels (time-series of cross-sections) of all sorts. For example, Kenneth French and I use the approach to examine issues in corporate finance (Fama and French 1998, 2002). In applications in which the dependent variable in the regression is asset returns, autocorrelation of the period-by-period regression slopes (which are portfolio returns) is not a problem. When autocorrelation of the slopes is a problem, as is more likely in other applications, correcting the standard errors of the average slopes is straightforward.

Outside of finance, research in economics that uses panel regressions has slowly come to acknowledge that residual covariance and autocorrelation are pervasive problems. Robust regression “clustering” techniques are now available (for example, Thompson 2011). The Fama-MacBeth approach is a simple alternative.

B. The Problems of the CAPM

The evidence in Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) is generally favorable to the CAPM, or at least to Black’s (1972) version of the CAPM in which there is no risk-free security. The golden age of the model is, however, brief. In the 1980s, violations, labeled anomalies, begin to surface. Banz (1981) finds that market $\beta$ does not fully explain the higher average returns of small (low market capitalization) stocks. Basu (1983) finds that the positive relation between the earning-price ratio ($E/P$) and average return is left unexplained by $\beta$. Rosenberg, Reid, and Lanstein (1985) find a positive relation between average stock return and the book-to-market equity ratio ($B/M$) that
is missed by the CAPM. Bhandari (1988) documents a similar result for market leverage (the ratio of debt to the market value of equity, \( D/M \)). As noted earlier, Ball (1978) argues that variables like size, \( E/P \), \( B/M \), and \( D/M \) are natural candidates to expose the failures of asset pricing models as explanations of expected returns since all these variables use the stock price, which, given expected dividends, is inversely related to the expected stock return.

Viewed one at a time in the papers that discovered them, the CAPM anomalies seemed like curiosity items that show that the CAPM is just a model and can’t be expected to explain the entire cross-section of expected stock returns. In updated tests, Fama and French (1992) examine all the common anomalies. Apparently, seeing all the negative evidence in one place leads readers to accept our conclusion that the CAPM just doesn’t work. The model is an elegantly simple and intuitively appealing tour de force that lays the foundations of asset pricing theory, but its major prediction that market \( \beta \) suffices to explain the cross-section of expected returns seems to be violated in many ways.

In terms of citations, Fama and French (1992) is high on the Journal of Finance all-time hit list. Its impact is somewhat surprising since there is little new in the paper, aside from a clear statement of the implications of the accumulated empirical problems of the CAPM.

C. The Three-Factor Model

An asset pricing model can only be replaced by a model that provides a better description of average returns. The three-factor model of Fama and French (1993) addresses this challenge. The model’s expected return equation is,

\[
E(R_{it}) - R_{ft} = b_1[E(R_{Mt}) - R_{ft}] + s_1E(SMB_i) + h_1E(HML_i).
\]

The time-series regression used to test the model is,

\[
R_{it} - R_{ft} = a_1 + b_1(R_{Mt} - R_{ft}) + s_1SMB_i + h_1HML_i + e_{it}.
\]

In these equations \( R_{it} \) is the return on security or portfolio \( i \) for period \( t \), \( R_{ft} \) is the risk-free return, \( R_{Mt} \) is the return on the value-weight (VW) market portfolio, \( SMB_i \) is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, \( HML_i \) is the difference between the returns on diversified portfolios of high and low \( B/M \) stocks, and \( e_{it} \) is a zero-mean residual. The three-factor model (13) says that the sensitivities \( b_1, s_1, \) and \( h_1 \) to the portfolio returns in (14) capture all variation in expected returns, so the true value of the intercept \( a_1 \) in (14) is zero for all securities and portfolios \( i \).
The three-factor model is an empirical asset pricing model. Standard asset pricing models work forward from assumptions about investor tastes and portfolio opportunities to predictions about how risk should be measured and the relation between risk and expected return. Empirical asset pricing models work backward. They take as given the patterns in average returns, and propose models to capture them. The three-factor model is designed to capture the relation between average return and size (market capitalization) and the relation between average return and price ratios like the book-to-market ratio, which were the two well-known patterns in average returns at the time of our 1993 paper.

To place the three-factor model in the rational asset pricing literature, Fama and French (1993) propose (13) as the expected return equation for a version of Merton’s (1973a) ICAPM in which up to two unspecified state variables lead to special risk premiums that are not captured by the market factor. In this view, size and $B/M$ are not themselves state variables, and $SMB$ and $HML$ are not portfolios that mimic state variables. Instead, in the spirit of Fama (1996), the factors are just diversified portfolios that provide different combinations of covariances with the unknown state variables. And the zero intercepts hypothesis for (14) implies that the market portfolio, the risk-free asset, $SMB$ and $HML$ span (can be used to generate) the relevant multifactor efficient set. In this scenario, (13) is an empirical asset pricing model that allows us to capture the expected return effects of state variables without naming them.

There is another more agnostic interpretation of the zero-intercepts hypothesis for (14). With risk-free borrowing and lending, there is one “tangency” portfolio of risky assets that is the risky component of all the mean-variance-efficient portfolios of Markowitz (1952). If the tangency portfolio can be expressed as a portfolio of the risk-free asset, the market portfolio, $SMB$ and $HML$, the analysis in Huberman and Kandel (1987) implies that the intercept in (14) is zero for all assets. This view of the three-factor model covers the ICAPM interpretation of Fama and French (1993) and the behavioral stories discussed later.

Kenneth French and I have many papers that address the empirical robustness of the three-factor model and the size and $B/M$ patterns in average returns the model is designed to explain. For example, to examine whether the size and $B/M$ patterns in average returns observed for the post-1962 period in Fama and French (1992) are the chance result of data dredging, Davis, Fama, and French (2000) extend the tests back to 1927, and Fama and French (1998, 2012) examine international data. The results are similar to those in Fama and French (1992, 1993). Fama and French (1996, 2008) examine whether the three-factor model can explain the anomalies that cause problems for the CAPM. The three-factor model does well on the anomalies associated with size, sales growth, and
various price ratios, but it is just a model and it fails to absorb other anomalies. Most prominent is the momentum in short-term returns documented by Jegadeesh and Titman (1993), which is a problem for all asset pricing models that do not add exposure to momentum as an explanatory factor, and which in my view is the biggest challenge to market efficiency.

After 1993, empirical research that uses an asset pricing model routinely includes the three-factor model among the alternatives. When the issue is the performance of a proposed new asset pricing model, victory is declared if the model comes somewhat close to explaining as much of the cross-section of average returns as the three-factor model. Research on the performance of managed portfolios (for example, mutual funds) routinely uses the intercepts (“alphas”) produced by (14), often augmented with a momentum factor (for example, Carhart 1997, and more recently Kosowski et al. 2006 or Fama and French 2010).

A long time passed before the implications of the work on market efficiency for portfolio choice had an impact on investment practice. Even today, active managers (who propose to invest in undervalued securities) attract far more funds than passive managers (who buy market portfolios or whole segments of the market). This is puzzling, given the high fees of active managers and four decades of evidence (from Jensen 1968 to Fama and French 2010) that active management is a bad deal for investors.

In contrast, the work on the empirical problems in the CAPM model for expected returns, culminating in Fama and French (1992, 1993), had an immediate impact on investment practice. It quickly became common to characterize professionally managed portfolios in terms of size and value (high \( B/M \)) or growth (low \( B/M \)) tilts. And it quickly became common to use the regression slopes from the three-factor model to characterize the tilts and to use the intercept to measure abnormal average returns (alpha).

There is longstanding controversy about the source of the size and especially the value premium in average returns. As noted above, Fama and French (1993, 1996) propose the three-factor model as a multifactor version of Merton’s (1973a) ICAPM. The high volatility of the SMB and HML returns is consistent with this view. The open question is: what are the underlying state variables that lead to variation in expected returns missed by the CAPM market \( \beta \)? There is a literature that proposes answers to this question, but the evidence so far is unconvincing.

The chief competitor to our ICAPM risk story for the value premium is the overreaction hypothesis of DeBondt and Thaler (1987) and Lakonishok, Shleifer, and Vishny (1994). They postulate that market prices overreact to the recent good times of growth stocks and the bad times of value stocks. Subsequent price
corrections produce the value premium (high average returns of value stocks relative to growth stocks). The weakness of this view is the presumption that investors never learn about their behavioral biases, which is necessary to explain the persistence of the value premium. Moreover, Fama and French (1995) find that the high average returns of value stocks and the low average returns of growth stocks persist for at least five years after stocks are allocated to value and growth portfolios, which seems rather long to be attributed to correction of irrational prices.

Asset pricing models typically assume that portfolio decisions depend only on properties of the return distributions of assets and portfolios. Another possibility, suggested by Fama and French (2007) and related to the stories in Daniel and Titman (1997) and Barberis and Shleifer (2003), is that tastes for other characteristics of assets play a role. ("Socially responsible investing" is an example.) Perhaps many investors get utility from holding growth (low $B/M$) stocks, which tend to be profitable fast-growing firms, and they are averse to value stocks, which tend to be relatively unprofitable with few growth opportunities. If such tastes persist, they can have persistent effects on asset prices and expected returns, as long as they don't lead to arbitrage opportunities. This is a behavioral story, but it is not about irrational behavior. In economics, we take tastes as given and make no judgments about them.

To what extent is the value premium in expected stock returns due to ICAPM state variable risks, investor overreaction, or tastes for assets as consumption goods? We don't know. An agnostic view of the three-factor model that doesn't require a choice among stories is that the model uses empirical regularities observed in many markets to find portfolios that together span the mean-variance-efficient set of Markowitz (1952). The analysis in Huberman and Kandel (1987) then implies that the model can be used to describe expected returns on all assets.

**Conclusions**

In my view, finance is the most successful branch of economics in terms of rich theory, extensive empirical tests, and penetration of the theory and evidence into other areas of economics and real-world applications. Markowitz’ (1952, 1959) portfolio model is widely used by professional portfolio managers. The portfolio model is the foundation of the CAPM of Sharpe (1964) and Lintner (1965), and it gets a multifactor extension in Merton (1973a). The CAPM is one of the most extensively tested models in economics, it is well-known to students in areas of economics other than finance, and it is widely used by practitioners.
The options pricing model of Black and Scholes (1973) and Merton (1973b) is a must for students in all areas of economics, and it is the foundation for a huge derivatives industry. However one judges market efficiency, it has motivated a massive body of empirical work that has enhanced our understanding of markets, and like it or not, professional money managers have to address its challenges. Its sibling, rational expectations, first exposited by Muth (1961), has had a similar run in macroeconomics. The three-factor model of Fama and French (1993) is arguably the most successful asset pricing model in empirical tests to date, it can't be avoided in tests of competing asset pricing models, and it is a handy tool that has shaped the thinking of practitioners. Can any other branch of economics claim similar academic and applied impact?

Acknowledgments


Literature Cited


Two Pillars of Asset Pricing


Portrait photo of Eugene Fama by photographer Alexander Mahmoud.