Two Pillars of Asset Pricing

By Eugene F. Fama

The Nobel Foundation asks that the Nobel 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.

I. Efficient Capital Markets

A. Early Work

The year 1962 was a propitious time for PhD 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 labelled 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 inefficient market or a bad model of market equilibrium. This is the joint hypothesis problem emphasized in Fama (1970).

A bit of notation makes the point precise. Suppose time is discrete, and $P_{t+1}$ is the vector of payoffs at time $t+1$ (prices plus dividends and interest payments) 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, suppose $f(P_{t+1} | \Theta_t)$ 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

\begin{equation}
\tag{1}
f(P_{t+1} | \Theta_{tm}) = f(P_{t+1} | \Theta_t).
\end{equation}

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 implied by time $t$ prices and $f(P_{t+1} | \Theta_t)$, the market efficiency condition is

\begin{equation}
\tag{2}
E(R_{t+1} | \Theta_{tm}) = E(R_{t+1} | \Theta_t).
\end{equation}

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 primitive—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,

\begin{equation}
\tag{3}
E(R_{t+1} | \Theta_{tm}) = E(R).
\end{equation}

If the market is efficient so that (2) holds, then

\begin{equation}
\tag{4}
E(R_{t+1} | \Theta_t) = E(R).
\end{equation}

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 on 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 et al. (1969) extend the tests to the adjustment of stock prices to announcements of corporate events. In Fama et al. 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,

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

In this regression, \( R_{it} \) is the return on stock \( i \) for month \( t \), \( R_{M_t} \) 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 percent 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 become 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,

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

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,

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

The expected inflation rate estimated in this way captures all the information used to set the interest rate. On 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 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 US 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 $t+1$,

$$R_{t+1} = a + bi_{t+1} + \varepsilon_{t+1}. \tag{9}$$

The tests say that monthly, quarterly, and semiannual nominal returns on longer-term bonds and real estate compensate for monthly, quarterly, and semiannual 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 use 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; 2006; 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 $t+T$, $F_{t,t+T}$, contains the expected term premium, $E(R_{t+1}) - S_{t+1}$, 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,

$$RT_{t+1} - S_{t+1} = a_1 + b_1(F_{t,t+T} - S_{t+1}) + e_{1,t+1}, \tag{10}$$

$$S_{t+T} - S_{t+1} = a_2 + b_2(F_{t,t+T} - S_{t+1}) + e_{2,t+T}. \tag{11}$$
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 2006).

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 through 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 semiannual expected stock returns using predetermined monthly, quarterly, and semiannual 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 (2011) 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 US 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 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 one needs to evaluate a forecaster’s entire track record, and, more important, the track records of all forecasters we might have chosen ex ante.

The absence of evidence that stock market price declines are predictable seems sufficient to conclude 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 US 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.

In percent terms, the largest five price declines in Figure 2 are (i) August 1929 to June 1932, (ii) October 2007 to February 2009, (iii) February 1937 to March 1938, (iv) August 2000 to September 2002, and (v) 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 Allan 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 US 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 percent 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 percent. The value to homeowners from housing services during the almost nine years from July 2003 to March 2012 surely exceeds 6.7 percent of July 2003 home values. Moreover, on the last sample date, October 2013, the real estate index, at 165.91, is 16 percent 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.

II. 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 of Sharpe (1964) and Lintner (1965). In this model market $\beta$, 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 $\beta$s.
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 $\beta$s and other variables. Black et al. (1972) criticize this approach because it produces estimates of the slope for $\beta$ (the premium in expected returns per unit of $\beta$) 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 $\beta$s 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 $\beta$ 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 $\beta$s, their (logged) market capitalizations at the beginning of month $t$, $MC_{i,t-1}$, and their book-to-market equity ratios, $BM_{i,t-1}$,

\begin{equation}
R_{it} = a + a_{1t} b_i + a_{2t} MC_{i,t-1} + a_{3t} BM_{i,t-1} + e_{it}.
\end{equation}

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 et al. (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 et al. (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

\begin{equation}
E(R_{it}) - R_{ft} = b_i[E(R_{Mt}) - R_{ft}] + s_i E(SMB_t) + h_i E(HML_t).
\end{equation}

The time-series regression used to test the model is

\begin{equation}
R_{it} - R_{ft} = a_i + b_i(R_{Mt} - R_{ft}) + s_i SMB_t + h_i HML_t + e_{it}.
\end{equation}

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, $SMB_t$ is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, $HML_t$ 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 $b_i, s_i,$ and $h_i$ to the portfolio returns in (14) capture all variation in expected returns, so the true value of the intercept $a_i$ 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, $SMB$ 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 et al. (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 long-standing 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 et al. (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.

III. 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’s (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?
REFERENCES


Speculative Asset Prices

By Robert J. Shiller

I will start this lecture with some general thoughts on the determinants of long-term asset prices such as stock prices or home prices: what, ultimately, drives these prices to change as they do from time to time and how can we interpret these changes? I will consider the discourse in the profession about the role of rationality in the formation of these prices and the trend toward behavioral finance and, more broadly, behavioral economics, the growing acceptance of the importance of alternative psychological, sociological, and epidemiological factors as affecting prices. I will focus on the statistical methods that allow us to learn about the sources of price volatility in the stock market and the housing market, and evidence that has led to the behavioral finance revolution in financial thought in recent decades.

The broader purpose here is to appreciate the promise of financial technology. There is a great deal of popular skepticism about financial institutions afoot these days, after the financial and economic crisis that has dragged on ever since the severest days in 2008. I want to consider the possibilities for the future of finance in general terms, rather than focusing on current stopgap measures to deal opportunistically with symptoms of our current economic crisis. The talk about the rationality of markets is a precursor to this talk of financial technology, for it underpins our notions of the possibilities that new technology offers.

I will conclude that the markets have already been “human-factors-engineered” to function remarkably well, and that as we improve our understanding of the kind of psychology that leads to bubbles and related problems, we can further innovate to improve the functioning of these markets.

I. Price Volatility, Rational Expectations, and Bubbles

The history of thought in financial markets has shown a surprising lack of consensus about a very fundamental question: what ultimately causes all those fluctuations in the price of speculative assets like corporate stocks, commodities, or real estate? One might think that so basic a question would have long ago been confidently answered. But the answer to this question is not so easily found.
At the same time, there has been an equally widespread acceptance in other quarters of the idea that markets are substantially driven by psychology. Indeed, since 1991 Richard Thaler and I have been directors of the National Bureau of Economic Research program in behavioral economics, which has featured hundreds of papers that seem mostly at odds with a general sense of rationality in the markets.²

The term “speculative bubble” is often used and applied carelessly. The word “bubble” first became popular at the time of the Mississippi Bubble in European stock markets that came to an end in 1720, a time often mentioned as one of craziness, but whether that period is best described as one of wild irrationality still remains controversial (see Garber 2000; Goetzmann et al. 2013). I would say that a speculative bubble is a peculiar kind of fad or social epidemic that is regularly seen in speculative markets; not a wild orgy of delusions but a natural consequence of the principles of social psychology coupled with imperfect news media and information channels. In the second edition of my book Irrational Exuberance I offered a definition of bubble that I think represents the term’s best use:

A situation in which news of price increases spurs investor enthusiasm which spreads by psychological contagion from person to person, in the process amplifying stories that might justify the price increases and bringing in a larger and larger class of investors who, despite doubts about the real value of an investment, are drawn to it partly through envy of others’ successes and partly through a gambler’s excitement.

At the center of my definition of the bubble are the epidemic spread, the emotions of investors, and the nature of the news and information media. Bubbles are not, in my mind, about craziness of investors. They are rather about how investors are buffeted en masse from one superficially plausible theory about conventional valuation to another. One thinks of how a good debater can take either side of many disputes, and, if the debater on the other side has weak skills, can substantially convince the audience of either side. College debate teams demonstrate this phenomenon regularly, and they do it by suppressing certain facts and amplifying and embellishing others. In the case of bubbles, the sides are changed from time to time by the feedback of price changes, at the proliferation caused by price increases of reminders of basic facts that a debater might use to defend the bubble. And the news media are even better at presenting cases than are typical college debaters.

Investing ideas can spread like epidemics. Economists traditionally have not shown much interest in epidemiology, sociology, social psychology, or communications and journalism, and it takes some effort for them to consider such alien academic traditions.

There is a troublesome split between efficient markets enthusiasts (who believe that market prices incorporate accurately all public information and so doubt that bubbles even exist) and those who believe in behavioral finance (who tend to believe that bubbles and other such contradictions to efficient markets can be understood only with reference to other social sciences such as psychology). I suspect that some of the apparent split is illusory, deriving from the problem that there is not a widely

accepted definition of the term “bubble.” The metaphor might suggest that speculative bubbles always burst suddenly and irrevocably, as soap bubbles seem to do, without exception. That would be silly, for history does not generally support the catastrophic burst notion. Though the abrupt ends of stock market booms in 1929, 2000, and 2007 might seem consonant with such a metaphor, these booms were reflated again before long (1933–1937, 2003–2007, and 2009–present, respectively).

I think that the eventuality of a sudden irrevocable burst is not essential to the general term speculative bubble as the phrase is appropriately used. The metaphor may be misleading: it suggests more drama than there in fact is, imparting a sense of uniqueness to current events, which might help explain the popularity of the term by news reporters vying for the attention of readers. Just as reporters like to stir people up by reporting that an index has hit another record high (disregarding the fact that record highs occur quite often, especially since reporters hardly ever correct for inflation), so too they like to suggest the possibility of a collapse in the offing that will be remembered many years later.

I sometimes wish we had a different metaphor. One might consider substituting the term “wind trade,” Dutch Windhandel, a term that was used during the Tulipmania, the famous boom and bust in tulip prices in the early 1600s. The reference to trading mere air seems more apt than the evocation of a fragile bubble.

Curiously, in his Nobel Lecture in Medicine during the 2013 Nobel Week in Stockholm, James E. Rothman (2013) involved soap bubbles too, for their analogy to the cell vesicles that were the focus of his Nobel Prize research. He showed a movie of two soap bubbles being pressed together, and, surprisingly to most of us, they did not burst but merged into a single larger bubble. That’s analogous to what cell vesicles can do, he said. It led me to wonder whether we could say that the stock market bubble and the housing bubble of the early 2000s somehow merged into a larger bubble that burst around 2008, touching off widespread financial crisis. Imaginative thinking is fun, and maybe even inspirational, but we cannot let the bubble metaphor, or any simple metaphor, guide our models beyond the very beginnings, for any metaphor will break down if we carried it to its absurd conclusions.

A. Efficient Markets Theory

From the very beginning, in his 1964 PhD dissertation, written under the supervision of Merton Miller and Harry Roberts, Eugene Fama found that stock prices are not very forecastable. He found then that the average correlation coefficient between successive day’s log price changes over the 30 Dow Jones Industrial Average stocks between 1957 and 1962 was only 0.03, which he described as “probably unimportant for both the statistician and the investor.”3 The same year saw the appearance of Paul Cootner’s The Random Character of Stock Market Prices, which reached similar conclusions about market efficiency.

The “efficient markets theory,” widely attributed to Fama and the academic work that he stimulated, maintains that prices have a rational basis in terms of fundamentals like the optimal forecast of earnings, or assessments of the standard deviation

3Fama (1964, Table 10 and p. 70).
of risk factors facing corporations. As the theory went, because they are rationally
determined, they are changed from day to day primarily by genuine news, which is
by its very nature essentially unforecastable. There was an efficient markets revolu-
tion in finance, propelled by Fama’s work. I was part of the movement then, less
than a decade later, with my PhD dissertation (1972) about the efficiency of the
long-term bond market.

B. Alternative Views and Forecastability of Returns

These conclusions came against a backdrop of public interest then in speculative
bubbles encouraged by the strong bull market in the United States: real stock prices
more than quadrupled in the 16 years from 1948 to 1964. John Kenneth Galbraith’s
best-selling 1954 book *The Great Crash 1929* described in literary terms the follies
of the boom of the 1920s and subsequent collapse, and concluded that “the chances
for a recurrence of a speculative orgy are rather good.”

His book was followed up by another popular book, Charles Poor Kindleberger’s
*Manias, Panics, and Crashes* (1978), which used a similar method of recounting
of human events laced with descriptions of human foolishness. Neither of them made
much use of academic research in psychology or sociology, writing many years
before the behavioral finance revolution, and so they came across to some as insub-
stantial. While both Galbraith and Kindleberger were respected academics, and the
stories in their books were often compelling, many felt that their works did not have
the scientific credibility of the careful data analysis that was widely taken to support
market efficiency, though, then again, they were provocative.

Ultimately, the question in reconciling the apparently conflicting views comes
down to that of constructing the right statistical tests. It turns out that the apparently
impressive evidence for market efficiency was not unimpeachable.

II. Expected Present Value Models and Excess Volatility

The simplest version of the efficient markets model—which maintains that stock
price movements can be interpreted entirely as reflecting information about future
payouts of dividends, and that hence there is never a good time or bad time to enter
the market—has, ever since the efficient markets revolution began, maintained a
powerful hold on scholarly imaginations as a worthy approximation to more com-
plex models. This form sets price equal to the expectation, conditional on publicly
available information at the time of the present value of future dividends discounted
at a constant rate through time:

\[ P_t = E_t \sum_{k=0}^{\infty} \frac{D_{t+k}}{(1 + r)^{k+1}} . \]

One way to test this efficient markets model is to regress the return between \( t \) and
\( t + 1, t = 1, \ldots, T \) onto information variables known at time \( t \), \( I_t, t = 1, \ldots, n \). Often,
these tests can be described approximately as tests of the “random walk hypothesis,” that price changes are purely random and unforecastable. One accepts the efficient markets model if the coefficients of the information variables used to forecast future returns or price changes are not significantly different from zero. Moreover, even if the model is rejected, if the proportion of variance in returns that is predicted is small, one concludes that the model is a good approximation to reality.

These tests, and various analogues of them, are the kinds of tests of market efficiency that abounded in the literature. But the power of such tests of perpetual unforecastability of returns against an alternative that represents the world as driven entirely by temporary fads and fashions—with no fundamental reason for any change in prices—can be very low since plausible such alternatives also imply that only a tiny fraction of month-to-month returns is forecastable (Shiller 1984, 1989; Summers 1986).

Many tests of market efficiency use daily observations of prices, and because the observations come so frequently, there may be many hundreds of observations, even if the span of the data is only a few decades. There is a tendency for many people to think that hundreds of observations must be a lot of data, but it is not necessarily a lot of data from the standpoint of distinguishing an efficient markets model from a relevant alternative.

We might, for example, be trying to determine whether some price time-series data is a random walk as against the alternative of a continuous-time first-order autoregressive process. In the former, whether prices are too high or too low has no ability to predict future changes. In the latter, when prices are too high relative to the mean they should tend eventually to fall (a sort of bursting of the bubble, though not a sudden catastrophic one). But, tests may have very little power to distinguish the two models, if the autoregressive parameter is close enough to one, even with a large number of observations, even with day-to-day or minute-to-minute observations. With a fixed span of data, increasing the frequency of observation, even to the limit of continuous observation, does not bring power to one (Shiller and Perron 1985; Phillips and Perron 1988).

The Scientific Background for the 2013 Nobel Prize in Economics (Economic Sciences Prize Committee of the Royal Swedish Academy of Sciences 2013) emphasized the results of this year’s laureates as confirming that there is better forecastability (in terms of $R$-squared) of speculative asset returns for longer time horizons. This accords with long-standing advice that investors should be patient, that they cannot expect to see solid returns over short time intervals. But this is just the opposite of what one would expect in weather forecasting, where experts can forecast tomorrow’s temperature fairly well but certainly cannot forecast accurately a year into the future.

It is easy to see why short-term forecastability of price changes in investable assets should in some sense be unlikely: if investment returns were substantially forecastable from day to day, it would be too easy to get rich in a year or so by trading on these forecasts, and we know it cannot be easy to make a lot of money trading. This notion was formalized in a continuous-time framework by Sims (1984), who defined “instantaneous unpredictability” of a speculative asset price by the requirement that

\footnote{In continuous time, we are speaking of distinguishing a Wiener process from an Ornstein-Uhlenbeck process.}
the $R$-squared of the prediction from time $t$ to time $t+s$ goes to zero as $s$ goes to zero. He showed under certain regularity conditions that if prices are not instantaneously unpredictable, then simple rapid-trading schemes could achieve unbounded profits, which of course cannot match reality.

Taking these primal reasons to doubt that returns are forecastable over short horizons into account, the low $R$-squared in many tests of short-run market efficiency are neither surprising nor interesting. The tests tell us only the obvious, and do not tell us about the rationality of markets beyond that people are not missing easy opportunities to get rich very fast.

I proposed that an alternative class of tests, based on the estimated volatility of returns—tests for “excess volatility”—would have more power against the important alternatives to efficient markets theory: first for the bond market, rejecting the expectations model of the term structure of interest rates with US and UK data (Shiller 1979), and then rejecting the simplest efficient markets model for the US stock market (Shiller 1981a). Independent work by Kenneth Singleton (1980) used a variance bounds test to reject the expectations model of the term structure of interest rates with US data, and Stephen LeRoy and Richard Porter (1981) rejected the simple efficient markets theory for the US stock market. Variance bounds tests were also used to test consumption-discount-based efficient markets models (Shiller 1982; Hansen and Jagannathan 1991). Efficient markets models also imply bounds on the covariance between asset prices (Beltratti and Shiller 1993).

These tests may be more powerful than regression tests of the basic efficient markets notions against important alternatives. It is true that under the conventional assumptions of the regression model the usual $t$-test for the coefficient of a forecasting variable in a regression with excess return as the dependent variable has well-known optimality properties. But 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. I showed (1981b) that if we broaden the maintained hypothesis for this condition, then a regression test is not optimal. In fact, under certain extreme assumptions about data alignment, a simple variance ratio test, instead of a regression test, may be uniformly most powerful.

Another kind of test of market efficiency is the event study, which is an analysis of the effects of a specified event (such as a stock split) on the price of an asset in the days before and after the event, taking many different examples of a kind of event and showing the average price performance. It is analogous to a test of the significance of coefficients in a regression of a panel of time series of daily returns of many stocks on a dummy variable representing the day of a certain kind of event and on dummies representing the days after the event became public. The test of market efficiency is a test for significance of the coefficients of the dummies corresponding to the event day and the days after the event.

---

6 The volatility tests were partly inspired by work Jeremy Siegel did (1977) which involved calculation of ex post rational price series.

7 Regression tests have pitfalls as well as when ratios involving price are used in regressions explaining future change in price, creating an endogenous variables problem. See Campbell and Shiller (1989).

8 John Cochrane, in his review of my volatility tests (1991, 1992), stressed a sense in which there is an equivalence of volatility tests and regression tests. But this is about the equivalence of null hypotheses, not equivalence of test power. Cochrane later followed this up with a paper (2007) recognizing the importance of the terminal condition; see also Lewellen (2004) and Campbell and Yogo (2006).
to days after the event. The first event study in the academic literature has been taken to be Dolley (1933), but, as the Scientific Background for the 2013 Nobel Prize in Economics notes, it was not until the impressive 1969 paper by Eugene Fama, Lawrence Fisher, Michael Jensen, and Richard Roll that showed that, conditioning on an event, one tends to see a lack of any consistent and important further price response after the event is public knowledge. Dolley in his 1933 article was immersed in all the details of stock splits, and of course did not mention efficient markets theory. Fama, Fisher, Jensen, and Roll, instead, showed evidence for this newly developed and expanded theory, evidence that could be seen visually impressively in a plot of stock returns before and after the event.

But, again, the efficient markets tests, which are essentially the same as regression tests, do not have the power to tell us whether there are also bubbles affecting prices, or even whether the major component of stock price movements comes from bubbles. The variance bounds test rejections of market efficiency could not be dismissed as correct but unimportant, as were the inefficiencies that the efficient markets literature had discovered, for they suggested that most of the variability of the aggregate stock market was not explainable as related to information about future fundamentals. Critics of the variance bounds tests became abundant, and I endeavored at first to answer some of them, answering Marsh and Merton (1986), and Kleidon (1986), with Shiller (1986, 1988). But the volume of the literature expanded beyond my abilities to respond, and significantly changed its direction as well. Sometimes the disagreements got abstract and seemed to raise deep issues about epistemology or the philosophy of logic. I must leave it to a broader professional consensus what is the outcome of this debate.

I collected my papers on the subject and summarized the literatures in my book Market Volatility (1989), at which point I largely abandoned my econometric work on excess volatility. Others continued the line of work, and much more has happened since.

A. Visual Portrayals of Excess Volatility and of the Stock Market as Forecaster

Just as event studies visually convinced many readers of some merits of efficient markets theory by showing event study plots, showing stock prices before and after an event, so, too, other simple plots seem to have been convincing in a different way that stock markets are really not so efficient. Figure 1 is an updated version of one that I showed in my 1981a paper, a third of a century ago, of the real level of the stock market since 1871, as well as the behavior through time of the actual present value of future real dividends discounted at a constant rate. The real stock price series is one published by Standard and Poor’s, called the S&P Composite (after 1957, the S&P 500) deflated by the US consumer price index.

The earlier version of this plot turned out to be the centerpiece of that paper, judging from the attention that others gave to it. Sometimes a simple plot seems to be more disturbing than a formal analysis. Looking at the data is like seeing a
photojournalist’s account of a historical event rather than reading a chronology: it is more immediate and invites intuitive comparisons.

To produce this figure, the present value of dividends for each date 1871–2013 was computed from the actual subsequent real dividends using a constant real discount rate \( r = 7.6 \) percent per annum, equal to the historical average real return on the market since 1871. For this figure, I was able to make use of the actual dividends, as published by Standard and Poor’s since 1926 (and extended back to 1871 by Alfred Cowles (1939) as I described in my book (1989)). We did not know dividends after 1979 when I published the original version of this figure, and we do not know at this writing of dividends after 2013.

For this lecture, in 2013 as I did in 1981, I made some simple assumptions about the as-yet-unseen future dividends, beyond 2013. This time I used a conventional dividend discount model, the Gordon Model, using the most recent 2013 S&P 500 real dividend as a base for forecasts of dividends after 2013 showing two alternative assumptions about those dividends. In one, I assumed that real dividends will grow forever from the last observed dividend, in 2013, at the same average growth rate as over the most recent ten years, 5.1 percent per year, which gives a 2013 value of 1292 for \( P^* \). In another, the calculations are the same but the growth rate of dividends after 2013 is taken as the geometric average growth rate over the last 30 years, 2.5 percent a year. This gives a 2013 value of 669 for \( P^* \). Both of these may be contrasted with real market values of the S&P 500 index over the year 2013 ranging from 1494 to 1802.\[10\]

\[10\]Jeremy Siegel (2005, 2008) has made the point that since the dividend payout rate for earnings has been trending down since World War II, dividend growth should be higher in the future than it was. If companies reinvest earnings rather than pay them out, they should have more dividends to pay in the future. The validity of this theory is not without doubters. Arnott and Asness (2003) point out that lower dividend payouts may reflect managers’ decision in the face of evidence that they have that earnings growth will be lower.
Should we take the latest 10 years real dividend growth as a guide to the future, rather than the last 30 years or some other interval? The ten-years data are more recent, but ten years is a short time historically speaking, and the years 2003 to 2013 were unusual, starting with the aftermath of the 2001 recession, and encompassing the biggest financial crisis, and government stimulus packages, since the Great Depression. Reasonable people will certainly find reasons to differ. Worse than that, there is no objective way to forecast dividends out for decades, which is why I showed both here, as a crude indication of uncertainty today about future dividends and why it is hard to imagine that the market somehow “knows” the correct optimal forecast.

The point of showing the two different $P^*$ series is that, clearly, there is substantial uncertainty about the present value of dividends after 2013, but there is not so much variability from year to year, as seen today, about the present value of subsequent dividends for earlier years. For earlier years, say before 1980, 2013 is so far in the future and is discounted so heavily that over a wide range of possible 2013 dividend values there is not much difference in $P^*$.

The striking fact is that by either assumption the present value of dividends (on the log scale used in the figure) looks pretty much like a steady exponential growth line, while the stock market oscillates a great deal around it. I asked in 1981: if, as efficient markets theory asserts, the actual price is the optimal forecast as of any date of the present value as of that date, why is the stock market so volatile?

Different people have different reactions to this figure, but a common reaction is that the efficient markets model $P_t = E_t(P^*_t)$ looks implausible here. Why is price jumping around so much if the value it is tracking is just such a simple trend? It is not that $P_t$ should always look smoother through time than $P^*_t$, for it is consistent with the model that there can be sudden shifts in price when there is important new information about subtle changes in trend. But it would seem that important new information should be something that occurs only rarely, given the smooth nature of dividends.

To see the problem for efficient markets here, imagine that the series labeled $P^*_t$ is not price but air temperature, and that $P_t$ is a weather forecaster’s forecast of the temperature for that day $t$. We might be inclined to label this weather forecaster as insane. Even though in the stock market there isn’t immediate feedback to the forecaster about forecast errors, still a forecaster should avoid adjusting forecasts up and down frequently, unless there is actual new information, and clearly there wasn’t, not information about something that actually happened in stock market history.

One very basic thing that is learned from this figure is that the model that people essentially know the future, a model that is often suggested as an approximation, is wildly wrong in all periods. Sometimes people have suggested that the low stock prices seen in the Great Depression of the 1930s were justified because people rationally saw the damage to future real dividends caused by the Depression. But, in fact, at the worst of the stock market depression, in 1932, subsequent dividends just weren’t low enough for long enough to depress $P^*_{1932}$ by much at all. Nothing has ever deflected real dividends for very long from a long-run growth trend of a couple percent a year.

In my original paper (1981), I detrended the data (as is shown in a reproduction of that plot in the Scientific Background (2013) shown on the Nobel Foundation website), thinking that it is reasonable to assume that people know the trend. Under that assumption, the efficient markets model implies that the variance around trend should be less for $P$ than for $P^*$, which is plainly not the case in Figure 1. But, there
was a lot of negative reaction by critics of my paper to the assumption that the trend is essentially known.

Generally, these criticisms held that there was always some reason to think that the path of dividends might eventually depart markedly from its historical growth path, and that investors were evaluating constant new information about that possibility, and that they were rational to do so even if the dividend growth path never deviated far for long from a trend. This assumes that all the fluctuations are because of genuine information about those “black swan” outlier events that might have happened during a period of more than a century but just didn’t happen. Some of the criticism had to do with the possibility that the dividend series might have a unit root, and so that the apparent smooth trend was just a chance outcome that might not be continued into the future.\footnote{Unit root problems pose potentially serious problems for financial econometrics. See Campbell and Shiller (1988a); Torous, Valkanov, and Yan (2004); Campbell and Yogo (2006); and Cochrane (2007). Campbell and Shiller (1988a) proposed log-differencing to recast excess volatility tests in more robust terms. Fama and French (1988) and Poterba and Summers (1988) showed tests of simple efficient markets models based on ratios of variance of returns of different horizons. West (1988) showed an inequality in terms of variances of innovations of prices and present values, which strengthened the evidence for excess volatility.}

The uncertainty about the present value of dividends after 2013 as shown in Figure 1 does highlight an important problem. At every point in history there must have been some such uncertainty about future dividends. There are always factors or prominent theories that creative minds can bring up that would suggest a higher or lower rate of growth of dividends in the future.

For example, can we tell an efficient markets story why the stock market was so low in the Great Depression? The present value of actual future dividends was not particularly low in the Depression, but maybe people thought that they would be low, given the extant theories of the time. Or maybe they thought that the government would eventually nationalize the stock market without compensation. One might say that it would not be manifestly irrational, not crazy, to believe such stories. But why, then, do these stories come and go through time, causing the fluctuations in the market?

**B. Variations on the Present Value Model**

Of course, as we have noted, the basic notion of efficient markets does not necessarily require that discount rates are constant or that returns are not forecastable. A more general form of efficient markets would allow discount rates to depend on the time-varying one-period rate of interest:

\[
P_t = E_t(P_t^{*r}) = E_t \sum_{k=0}^{\infty} \prod_{j=0}^{k} \frac{1}{(1 + r_{t+j} + \varphi)} D_{t+k}.\]

Or, in a model proposed by LeRoy (1973) and Lucas (1978), and developed by Grossman and Shiller (1981) and Hansen and Singleton (1983), it could depend on consumption, using the marginal rate of substitution between consumption in successive periods as a discount rate:

\[
P_t = E_t(P_t^{*c}) = E_t \sum_{k=0}^{\infty} \prod_{j=0}^{k} M_{t+j} D_{t+k},\]
where \( M_t \) = marginal rate of substitution in consumption between \( t \) and \( t + 1 \), which is, assuming constant relative risk aversion \( A \), \( \rho(C_t/C_{t+1})^A \), and \( C_t \) is real per capita consumption at time \( t \).

Figure 2 shows the actual stock price in the US and the perfect foresight stock price corresponding to each of the three measures. Once again, the figure reveals that there is little correspondence between any of these measures of ex post rational price and actual stock price. People did not behave, in setting stock prices, as if they knew the future of these variables and reacted rationally to this knowledge. Moreover, if we assume that they did not actually have knowledge of the future, then one is led to wonder why the actual stock prices varied through time as much or more than the perfect foresight prices did.

There are continuing attempts to modify the consumption-based model to improve its fit (Campbell and Cochrane 1999 and Lars Hansen, in his Nobel Lecture (2013)) but not yet any model that could be set along side Figure 2 here as an inspiring vindication of efficient markets theory.

John Campbell and John Ammer (1993) did a variation decomposition (along lines developed in Campbell 1991) of unexpected excess returns using time series methods and US postwar data. The decomposition is based on the log linearization of the present value relation used in Campbell and Shiller (1988b). The time-\( t + 1 \) innovation \( E_t - E \) in the excess return over the risk-free rate \( e_t \) can be shown, with a terminal stationarity condition, as a tautology, to be the sum of three innovations:

\[
e_t + 1 - E_t = \left( E_t + 1 - E_t \right) \left\{ \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - \sum_{j=0}^{\infty} \rho^j r_{t+1+j} - \sum_{j=1}^{\infty} \rho^j e_{t+1+j} \right\}.
\]

Using this decomposition and a vector autoregressive model in difference form, with post-World War II stock market returns, Campbell and Ammer found that excess returns innovations have a standard deviation that is two or three times greater than the standard deviation of innovations in future dividend growth. Aggregate stock market fluctuations have therefore been dominated by fluctuations in predicted future returns, not by news about future dividends paid to investors.

C. Interpretations of Return Predictability

Sociologists have a possible interpretation of these results, an interpretation that reflects a body of thought that goes back over 100 years. The market fluctuates as the sweep of history produces different mindsets at different points of time, different zeitgeists. Durkheim (1893) spoke of the “collective consciousness,” that represents the shared beliefs, attitudes, and moral judgments that characterize a time. Halbwachs (1925) spoke of the “collective memory,” the set of facts that are widely

---

12 The parameter \( \varphi \) was estimated to make the average \( r_t + \varphi \) equal the average real return on the stock market 1871–2013. The parameter \( A \) was set at four and \( \rho \) at one. The one-year interest rate is pieced together from various sources as described in Shiller (1989, 2005) and real per capita consumption is from the US National Income and Product Accounts.

13 These results have been criticized by Goyal and Welch (2003, 2008), Chen and Zhao (2009), and Chen, Da, and Zhao (2013), and rebutted by Campbell, Polk, and Vuolteenaho (2010) and Engsted, Pedersen, and Tanggaard (2012).
remembered at any point of time, but that are forgotten eventually if word of mouth and active news media do not perpetuate their memory. News media tend to slant their stories toward ideas of current interest, rather than useful facts that readers no longer find interesting.\(^\text{14}\) Surely simple forgetting of past experiences affects popular judgments. How many people today could give any account of the financial panic of 1907, or of the housing boom of the late 1940s? One could stop anyone on the street in those times and get a ready account, now blank ignorance from almost everyone. When a bubble is building, the suppression of some facts and embellishment of other facts (just as with winning college debaters) occurs naturally through the decay of collective memory, when media and popular talk are no longer reinforcing memories of them, and the amplification of other facts through the stories generated by market events.

It is hardly plausible that speculative prices make effective use of all information about probabilities of future dividends. It is far more plausible that the aggregate stock market price changes reflect inconstant perceptions, changes that Keynes referred to with the term “animal spirits,” changes that infect the thinking even of most of the so-called “smart money” in the market. Keynes anticipated this in his 1921 *A Treatise on Probability*, which asserted that probabilities are not precisely measurable in the sense that decision theory supposes, that there are always ambiguities. He said that because of these fundamental ambiguities, there is, in financial transactions, inevitably an “element of caprice.”\(^\text{15}\) Critical decisions are made on impulse rather than calculation. One may have done calculations of


\(^{15}\) Keynes (1921, p. 23).
probabilities, but one usually does not fully believe one’s own calculations, and proceeds on gut feeling.

In an early behavioral finance paper of mine, “Stock Prices and Social Dynamics” (1984), I proposed yet another expected present value model for consideration as a model of stock prices, though it is one that we cannot plot back to 1871 as we did with the three expected present models shown and plotted above, because it depends on a time-varying factor that is not objectively quantifiable, at least for now. I have been attempting to measure a stock market factor like this with survey techniques, of individual and institutional investors, but only since 1989. There are other surveys of investor sentiment as well, but the results are hardly definitive. My surveys of individual and institutional investors starting in 1989,16 as well as my surveys with Karl E. Case of home buyers starting in 1988,17 are being continued by the Yale School of Management.

Thirty years ago I called this as yet unmeasured factor the “demand for stocks by ordinary investors,” but today let us call it animal spirits, \( A_t \). \( A_t \) represents the demand for stocks per share at time \( t \) everyone who is not smart money, people not really paying attention, not systematic, not engaged in research, buffeted by casually encountered information. They are certainly the majority of investors, and suppose, to take this model to an extreme, that their opinions reflect nothing more than changing fashions and fads and idle talk and overreaction to irrelevant news stories. \( A_t \) is likely to be sluggish through time (usually people don’t all change their naïve opinions en masse on a dime).

The core idea here was that there are also smart money investors, who are not subject to illusion, but have to be wary of investing in the stock market because not only are future dividends not known with certainty, but also because these ordinary investors are somewhat unpredictable and their erratic behavior could cause price changes that might produce losses in the market for the smart money if they invest too much in it. For these smart money investors, information is constantly coming in about the likely future values of \( A_t \) and, as with all genuinely new information, this new information is uncorrelated and unpredictable through time. I supposed the demand per share for stocks by the smart money equals their rationally expected excess return on the stock market over and above an alternative riskless return \( r \), which I take for simplicity to be constant through time, the difference divided by a constant risk factor \( \varphi \). The two demands, the demand of the ordinary investors plus the demand of the smart money, must add up to one for the markets to clear. Solving the resulting rational expectations model forward leaves us with our fourth present value model:18

\[
P_t = E_t(P_t^A) = E_t \left( \sum_{k=0}^{\infty} \frac{1}{(1 + r + \varphi)^{k+1}} (D_{t+k} + \varphi A_{t+k}) \right).
\]


18This is equation (3) in that paper, with slight changes in notation.
If $\varphi = 0$, smart money dominates, and this collapses to equation (1) above. As $\varphi$ goes to infinity, smart money drops out, it collapses to $P_t = A_t$, and ordinary investors completely determine the price. It is the intermediate case that is interesting. In this intermediate case, price may have low predictability from day to day or month to month, consistent with efficient markets theory, even if animal spirits dominate the broad movements in $P_t$. Long slow swings in $A_t$ may produce long slow swings in stock prices (perhaps the multiyear “bull” and “bear” markets) even though day-to-day movements in stock prices are nearly uncorrelated through time. The price is responding to news about animal spirits, not just news about future dividends. Event study tests, described above, testing market reaction over time to news about and subsequently reality of such events as stock splits, may come out as beautifully supporting efficient markets, for much of the effect of the event on both dividends and animal spirits will be incorporated into price as soon as the event becomes news to the smart money, not when the event actually happens.

There is another important argument widely used for efficient markets, the argument that a model like (4) with an intermediate $\varphi$ cannot represent a stable equilibrium because the smart money would get richer and richer and eventually take over the market, and $\varphi$ would go to zero. In fact this will not generally happen, for there is a natural recycling of investor abilities, the smart money people usually do not start out with a lot of money, and it takes them many years to acquire enough wealth to influence the market. Meanwhile they get old and retire, or they rationally lose interest in doing the work to pursue their advantage after they have acquired sufficient wealth to live on. The market will be efficient enough that advantages to beating the market are sufficiently small and uncertain and slow to repay one’s efforts that most smart people will devote their time to more personally meaningful things, like managing a company, getting a PhD in finance, or some other more enjoyable activity, leaving the market substantially to ordinary investors. Genuinely smart money investors cannot in their normal life cycle amass enough success experience to prove to ordinary investors that they can manage their money effectively: it takes too many years and there is too much fundamental uncertainty for them to be able to do that assuredly, and by the time they prove themselves they may have lost the will or ability to continue (Shiller 1984; Shleifer and Vishny 1997).

D. Individual Stocks

These conclusions about the aggregate stock market, however, do not carry over fully to individual stocks. Paul Samuelson has asserted that:

[The market is] micro efficient but macro inefficient. That is, individual stock price variations are dominated by actual new information about subsequent dividends, but aggregate stock market variations are dominated by bubbles.\(^{19}\)

Tuomo Vuolteenaho (2002), using methodology analogous to that of Campbell and Ammer, concluded that for individual stocks the variance of expected return news

\(^{19}\)Samuelson went on to say, “Modern markets show considerable micro efficiency (for the reason that the minority who spot aberrations from micro efficiency can make money from those occurrences and, in doing so, tend to wipe out
is approximately one-half of the variance of cash-flow news. For market-adjusted individual stock log returns (log return minus cross-sectional average log return), the variance of the expected return news is only one-fifth of the variance of cash-flow news. Thus, bubbles and their bursts cannot have more than a minority impact on the returns of individual stocks, and most of the variation in their returns comes from news about the future payouts the firms will make.

In a 2005 paper I did with Jeeman Jung, which looked at long-span datasets of stocks which had survived without significant capital changes for over half a century, we reached similar conclusions. To give a visual impression how well the efficient markets theory works for individual firms, we felt that we could display how successfully dividend growth could be predicted from the dividend-price ratio. Simple efficient markets suggest that firms with relatively low dividend price ratios should eventually, in future years, show higher dividend increases as a fraction of today’s price. To make such a visual diagram in such simple terms, we sought out long-lived firms (though such a procedure risks a selection bias).

We found all firms on the Center for Research in Security Prices (CRSP) tape that remained alive and for which there was uninterrupted data from 1926 to 1976. There were only 49 such firms, giving us 2,499 firm-year observations 1926–1976. Each point on the scatter in Figure 3 shows $$\sum_{k=0}^{24} \frac{\Delta D_{t+k}/P_t}{P_t^\gamma},$$ the present value of future changes in dividends for the next 25 years (measured in dollars, and discounted by the historical average stock market return) divided by current dollar price, against $$\frac{D_{t-1}}{P_t^\gamma},$$ the current dividend divided by current price. Efficient markets with constant discount rate, equation (1), implies, if there is not a problem with our truncation of the present value at 25 years, that a regression line through these points should have a slope of $-1.0$ and a constant term equal to the constant discount rate. In words, if markets are efficient then a high dividend price yield for a particular stock today occurs only if people have a real reason to expect dividends to decline, and so demand to be compensated today for that future loss if they are to hold the stock today. Similarly, low dividend yield stocks must be those for which there is genuine evidence that dividends will rise in the future, eventually compensating today’s investors for the low dividend return they are receiving.

The estimated slope of a line fitted through this scatter is $-0.5$, far from the ideal $-1.0$ but negative as expected. The dividend-price ratio predicts subsequent dividend changes in the right direction for these firms. Zero-dividend firms (which one can seestrung out along the vertical axis) tended to have appropriately high subsequent dividend growth relative to price. The right-most observation, which corresponds to the firm Schlumberger in 1931, a firm that had tried to maintain its dividend despite falling fortunes in the Great Depression, had a dividend payment that was 40 percent of its current price. People in the market then apparently figured out that the firm could not continue to pay such a dividend, that it would not be followed by any persistent inefficiencies). In no contradiction to the previous sentence, I had hypothesized considerable macro inefficiency, in the sense of long waves in the time series of aggregate indexes of security prices below and above various definitions of fundamental values.” From a private letter from Paul Samuelson to John Campbell and Robert Shiller.
another significant dividend for a very long time, and reflected that knowledge in
the approximately 40 percent dividend-price ratio. They were right to do this, as we
see after the fact. In individual firms there is sometimes a lot of action in the ratios,
and the action in fact often reflects real knowledge about future cash flows. That is
an example of the kind of idiosyncratic knowledge about individual firms that makes
the efficient markets model a useful approximation of reality for individual firms.20

E. Real Estate Prices

The market for real estate is larger in valuation than that of the entire stock mar-
et. According to the Financial Accounts of the United States, in 2013 the value
of real estate owned by households and nonprofit organizations was $21.6 trillion,
while their holdings of corporate equity shares, whether directly or indirectly, had a
market value of only $20.3 trillion.21

And yet, in the 1980s, when I first joined with Karl Case to do joint work on
real estate prices, we found that hardly any scholarly research had been done on
the efficiency of real estate markets. The state of knowledge about these markets
was abysmal. Under the influence then of a widely held presumption at that time
that all markets must be efficient, many economists, at least in their popular pro-
nouncements, seemed to assume that real estate markets must be efficient too. This
presumption appeared to us as quite probably wrong, based on anecdotal evidence

20 Ang and Bekaert (2007) conclude that the dividend yield’s ability to predict dividends is not robust over
sample periods or countries, but they do not include individual stock data in their study.
21 US Federal Reserve Board, Z.1, Financial Accounts of the United States, Table B.100 Balance Sheet
of Households and Nonprofit Organizations, and Table B.100.e Balance Sheet of Households and Nonprofit
Organizations with Equity Detail, December 9, 2013.
suggesting that real estate prices are not at all well approximated by a random walk, as is the case for stocks, but often tend to go in the same direction, whether up or down, again and again for years and years.

Case and I decided to try to test the efficiency of this market for single-family homes, but quickly discovered the importance of a stumbling block that had inhibited research: individual homes sell extremely infrequently, with interval between sales for individual homes measured not in minutes as with stocks but in years or decades. One cannot do any of the most popular tests of efficiency with such data. No runs tests or event studies would ever be possible with individual homes, and so tests of market efficiency would have to be based on indices.

There were some home price indices of sorts available then, but they had serious problems. There was a median sales price of existing homes, published by the National Association of Realtors, but it often appeared to jump around erratically. It was just the median price of whatever homes were selling at the moment and it was not controlled for any change in the composition of sales. Moreover, it appeared that different kinds of homes sold in different months. It had a very strong seasonal component, which we suspected arose because people who sold in the summer, in phase with the academic year and the job market, typically had bigger or higher-quality homes which had higher prices.

There was also at that time a “Price of New Homes Sold,” also called “Constant Quality Index,” produced by the US Census Bureau, that was a more sophisticated hedonic index, holding constant such things as square feet of floor space and number of bedrooms, but again it was obviously not trustworthy for testing market efficiency through time since it was based on different homes every quarter, whatever and wherever homes had just been built that quarter.

So Case and I constructed our own “repeat sales” home price index based on an inspiration of his (Case 1986) and then on a method we devised that inferred price changes only from the change in prices of individual existing homes (Case and Shiller 1987, 1989, 1990). We showed how a quarterly index could be computed even if homes sell much less frequently than quarterly. We discovered that Case’s inspiration was largely anticipated by Baily, Muth, and Nourse (1963), but we had a number of improvements, taking better account of heteroscedasticity. Later, I made the index arithmetic and value weighted, as are the most prominent stock price indices (Shiller 1991). With my former student Allan Weiss we founded Case Shiller Weiss, Inc., in 1991 and we were the first to produce repeat sales indices in real time for regular publication. We applied these indices to produce automated valuation models for single family homes (Shiller and Weiss 1999a). Our indices are now produced by CoreLogic, Inc., and the major indices managed by Standard and Poor’s Corporation.

A plot of our quarterly national index corrected for CPI inflation is shown in Figure 4 along with the Census Constant Quality Index, also converted to real terms. Simply producing these data and looking at a plot, as shown in Figure 4, yields some surprises. First of all, the home price data are generally extremely smooth through time, except for a small amount of seasonality. Home prices do indeed go through years of price increases and then years of price decreases. So, the random walk model of home price behavior is just not even close to being true for home prices (Case and Shiller 1988). Home prices might seem to be described as in
accordance with model (4) above with the parameter $\phi$ extremely large, so that the smart money, who might go in and out of the market quickly in response to news, is hardly a factor.

Second, while it was not apparent when we first computed these indices, it is clear from these data from today’s vantage point that there was a huge boom in home prices after 2000 that was not very visible from the Census Constant Quality Index. That boom was unprecedented in scope and magnitude in over 100 years of US history (Shiller 2005). Why is the boom and bust in home prices after 2000 so much more prominent in our repeat sales index? New homes are built where it is possible and profitable to build them, typically outside congested urban areas where price swings may be most pronounced, and so their level through time may be more nearly determined by simple construction costs. Thus, our data collection revealed not only market inefficiencies, but much bigger price swings as well.

The inefficiency that we documented in single family home prices must be related to market conditions, and so efficiency must be improvable with changes in market institutions. The inefficiency of the market for single family homes relative to that of the stock market must be partly traced to the relatively much higher cost of trading in that market. It is much more costly for professional traders to trade in and out of the market for single family homes to profit from predictable price movements. It is difficult to do short sales of overpriced individual homes. Buying and selling individual homes may not work well for professionals also because of high carrying costs, low rental income, moral hazard of the renters who have relatively little incentive to care for the property, and difficulty keeping up with all the local factors that might change the demand for individual houses, so that remote institutional investors would risk being picked off as ignorant losers. Some institutional investors
are in the news recently thinking they can survive and make money in this market. We will see if they succeed.

We thought that the market efficiency could be improved if an index of home prices could be made tradable (Case, Shiller, and Weiss 1993). Working with Standard and Poor’s, and with the people in our company MacroMarkets LLC, we helped the Chicago Mercantile Exchange with plans to set up futures markets based on our indices for ten US cities. These markets were launched in 2006, and are still trading today, albeit with nothing close to the volume of trade that we hoped to see.\(^{22}\) We hoped that the creation of these new markets would change the nature of prices in real estate markets, with price discovery that made the price of homes behave more like the random walk that efficient markets theory suggests.

Real estate markets remain wildly inefficient all over the world. We can only look forward to the day when liquid markets support more trade that might permit something rather closer to the efficient markets that theorists have expected.

To achieve such improvements in efficiency, in real estate markets, in stock markets, or in any speculative markets, it is most helpful to understand the causes of market inefficiency, and that requires serious study from the broad perspective afforded by an array of other sciences outside of economics.

### III. Behavioral Finance and Behavioral Economics

The behavioral economics revolution, which brings psychology and other social sciences into economics, saw its first beginnings in the 1980s, but did not attract public attention until the 1990s. Richard Thaler and I started our behavioral economics workshops at the National Bureau of Economic Research in 1991, and from the beginning behavioral finance played the dominant role.\(^{23}\) There are a number of surveys of the behavioral finance literature, notably Baker and Wurgler (2011), Barberis (2003), Shefrin (2008), Shiller (2003b), and Shleifer (2000).

The behavioral finance revolution seemed to take its beginnings from the evidence of market inefficiency that was by then starting to look significant. Once we acknowledge that the efficient markets theory has no special claim to priority for price determination, we can look more sympathetically to other factors to understand market fluctuations. The anomalies literature points indeed to some oddball factors as playing a role. Benos and Jochec (2013) showed that patriotism affects stock prices, in that US stocks with the words “America(n)” or “USA” in their names earn an abnormal return of 6 percent a year during wartime. Saunders (1993) found that the weather in New York affects stock prices. If such irrelevant things as these affect stock prices, it should be no surprise if more plausible but half-baked theories (about the central bank, fiscal policy, energy prices, the future of capitalism, and on and on) would also affect market prices.

---

\(^{22}\) See Fabozzi, Shiller, and Tunaru (2009). The market maker John Dolan has a website with up-to-date information about this market, http://homepricefutures.com. Our firm MacroMarkets LLC led by Allan Weiss and Sam Masucci also created paired long and short securities, MacroShares, with ticker symbols UMM (for up major metro) and DMM (for down major metro) based on the S&P/Case-Shiller Ten-City Index that traded on the New York Stock Exchange from August to December 2010.

Most stock market investors do not pay much attention to fundamental indicators of value. We might argue that their inattention is in some sense rational, since there is a cost to collecting information. Christopher Sims has devised a model of rational inattention (2003). But, it is hard to believe that their inattention is systematic and thoughtful.

A. Early History of Behavioral Finance

Behavioral finance until the 1980s was mostly relegated to the community of investment analysts who did not generally attract notice in academia, and who did not generally draw on research from the social sciences. There were however some gems from this period. Notable among these analysts were Benjamin Graham and David Dodd, who, in the 1940 edition of their book *Security Analysis*, based their investing method on their observations of “ignorance, of human greed, of mob psychology, of trading costs, of weighting of the dice by insiders and manipulators.”

Keynes gave a view of speculative markets that was ahead of its time. In his 1936 book *The General Theory of Employment, Interest and Money*, Keynes described speculative markets as akin to a newspaper competition he saw offered by a local newspaper to its readers. His metaphor is widely referred to as Keynes’s “beauty contest” theory of the stock market. Each reader was invited to submit from a page with 100 photos of pretty faces a list of the 6 of them he or she thought prettiest. The winner would be the one whose list most closely corresponded to the most popular faces among all the lists of six that readers sent in. Of course, to win this contest a rational person would not pick the faces that personally seem prettiest. Instead one should pick the six faces that one thinks others will think prettiest. Even better, one should pick the faces one thinks that others think that others think prettiest, or one should pick the faces one thinks that others think that others think that others think prettiest. The same is true with stock market investing. Keynes thought that “there are some, I believe, who practice the fourth, fifth and higher degrees,” further degrees of removal from reality than was embodied in equation (4) above. That is how speculative markets function, Keynes said. Active participants are trying to buy into their predictions of the conventional valuation of assets in the near future, not the true value.

A key Keynesian idea is that the valuation of long-term speculative assets is substantially a matter of convention, just as it is with judgments of facial beauty. Whatever price people generally have come to accept as the conventional value, and that is embedded in the collective consciousness, will stick as the true value for a long time, even if the actual returns fail for some time to live up to expectations. If an asset’s returns are carefully tabulated and disappoint for long enough, people will eventually learn to change their views, but it may take the better part of a lifetime. And many assets, such as owner-occupied homes, do not have unambiguously measured returns, and a mistaken “conventional valuation” based on a faulty popular theory can persist indefinitely. The presumed investment advantages of, say, living in an expensive land-intensive single-family home near a big city rather than renting

---

25 Keynes General Theory (1936, ch. 12, p. 156 (Harbinger 1965 edition)).
a cheaper and more convenient apartment in a high-rise there may just not exist, and most people will never figure that out.

Conventional valuation can be a very subtle phenomenon at any point of time, reflecting popular theories of the time that are perceived by many, who have never studied the theories, as reflecting professional wisdom. In a beauty contest, people have even less incentive to consider the validity of this wisdom, since they view it as substantially entrenched in others’ thinking. I am reminded, for example, of Modigliani and Cohn’s (1979) study showing that inflation-induced biases in conventional accounting practices caused a massive understatement of earnings, a study that allowed them to call roughly, within a few years, the historic bottom of the stock market in 1982. The absence of immediate reaction to their study was just the kind of thing one might expect to see in a beauty contest world, since no one expected anyone else to react much to their paper.

B. The Blossoming of Behavioral Finance after 1980

The idea that speculative prices are somehow uniquely authoritative, as the best possible judgment of true value, still has its popular appeal even today. But, it has lost its unique claim on the attention of economic theorists. Theoretical models of speculative markets that are analogous to Keynes’s beauty contest theory, that stress the expectation of reselling to other people who may have optimistic beliefs, have been offered by Harrison and Kreps (1978); Morris (1996); Scheinkman and Xiong (2003); Wu and Guo (2004); Hong, Scheinkman, and Xiong (2006); Allen, Morris, and Shin (2006); Hong and Sraer (2011); Kubler and Schmedders (2012); and Barberis et al. (2013). In addition, there are models that represent bubbles as related to leverage cycles tied in with heterogeneous beliefs: Fostel and Geanakoplos (2008), Geanakoplos (2009), Cao (2010), and He and Xiong (2012). Noise trader models (Kyle 1985; De Long et al. 1990; Campbell and Kyle 1993) have begun to replace models with all rational agents.

Moreover, there are models of financial markets that replace the assumption of rational expected-utility-maximizing agents with alternative models of human behavior, such as prospect theory (Kahneman and Tversky 1979, 2000). Prospect theory, which is a theory of human choice in the face of risk that is based on experimental evidence in the psychology laboratory, is not a theory of rationality in the traditional sense, for it recognizes violations of the basic axioms of rational behavior (Savage 1954). The human behavior prospect theory describes is vulnerable to the arbitrariness of psychological framing; insignificant changes in context or suggestion can produce profound differences in human behavior.

Barberis, Huang, and Santos (2001) showed that prospect theory with investors who derive direct utility from fluctuations in the value of their wealth can help explain the excess volatility of stock market returns. A “house money effect” can help make bubbles grow even bigger, in analogy to gamblers at casinos who, after they have won some money, become very risk tolerant with that money because they frame it as somebody else’s money that they can afford to lose (Thaler and Johnson 1990). Investors’ “narrow framing” (Barberis, Huang, and Thaler 2006) and the disposition to sell winners and hold losers (Shefrin and Statman 1984) can explain other evidence against efficient markets.
The field of psychology offers many other principles of human behavior that have been shown to be relevant for evaluating the efficient markets theory. For example, there is evidence that a general human tendency toward overconfidence causes investors to trade too much (Odean, 2000) and CEOs to squander internally generated funds on pet projects (Malmendier and Tate 2005). There is a tendency for investors to be overly distracted by news stories (Barber and Odean 2008) and to overreact to cash dividends (Shefrin and Statman 1984).

Financial theory has also advanced to allow us a better understanding of the effects of the ambiguity regarding probabilities, the fundamental difficulties in placing numerical values for probabilities, that Keynes spoke of (Bewley 2002; Bracha and Brown 2013).

Psychologists have documented a tendency for people to anchor their opinions in ambiguous situations on arbitrary signals that are psychologically salient even if they are obviously irrelevant (Tversky and Kahneman 1974).

Neuroscience has begun to understand how the human brain handles ambiguity. Hsu et al. (2005) and Huettel et al. (2006) use functional magnetic imaging to study brain reactions to situations with clear versus ambiguous probabilities. Huettel et al. concluded that “decision making under ambiguity does not represent a special, more complex case of risky decision making, instead, these two forms of uncertainty are supported by distinct mechanisms.”26 The rapid progress we are now seeing in neuroscience will likely yield new insights into the ambiguity, animal spirits, and caprice that Keynes and others since him have stressed.

### IV. Implications for Financial Innovation

The financial institutions that we have today are the product of centuries of experience with the volatility of speculative asset prices, with the important information discovery that these market prices can reveal, as well as the potential for erratic behavior in these markets.27 The reliability of these markets in revealing genuine information about fundamentals is not terrific, but it is certainly not negligible either, and the reliability might be improved through time with better financial institutions. Efficient markets should be considered a goal, not an established fact. The financial institutions that we have are the results of experimentation that has helped people design around this experience; the institutions we will have in the future depend on our continuing experimentation and redesign.

Like mechanical engineering, financial engineering should pay attention to human factors, to make devices that serve people well with full consideration of human talents and foibles. As this experience accumulates, with each successive financial crisis and each improvement in information technology, financial innovation can make these institutions work better for humankind.

For example, the very invention, centuries ago, of stock markets, has created an atmosphere for investing that, while it regularly produces the excesses of bubbles, creates an incentive for people to launch exciting new enterprises, to keep up to date

---

26 Huettel et al. (2006, p. 765).
27 For the long history of financial innovation, see Goetzmann and Rouwenhorst (2005).
on relevant information, and to protect themselves if they want from the inevitable risks of those very bubbles.

As Moss (2002) has chronicled, a general limited liability statute covering all stock market investors was not such an obviously good idea when the world’s first such law was passed in New York in 1811, but it turned out to be of fundamental importance for investors’ psychology. By clearly forbidding suing shareholders for a company’s sins, it limited the downside risk of investing to psychologically manageable proportions (no more worries that any one of your investments could explode and land you in debtors’ prison), and it permitted portfolio diversification to proceed without exhaustive investigation of each company’s management. The stock market became an exciting place, like a gambling casino, but tied to business reality rather than mere amusement, and it was a place where investors could diversify and limit their risks. It therefore was highly effective in attracting capital for enterprise.

More recently, people have been experimenting with other details of the stock markets, such as insider trader rules, risk retention rules, capital requirements, and other factors. These interact with human psychology in ways that can improve market functioning but whose effects cannot be accurately foretold from any received theory.

Much of my work has been involved in considering how both financial theory and human factors need to be considered in designing new financial structures. I have written a number of books devoted to this: Who’s Minding the Store? (1992), Macro Markets (1993), The New Financial Order (2003b), The Subprime Solution (2008), and Finance and the Good Society (2012). Most of the ideas I have expressed in those books are calls for experimentation, not finished ideas. The ideas I discussed are mostly as yet untested, and their final forms, if and when they ever do get implemented, perhaps in the distant future, and with far better information technology, are hard to see in advance.

The ideas in these books, and associated articles, are diverse, go in many directions, and have to be judged as beginnings of ideas. They may look awkward just as the earliest designs of aircraft did; their later incarnations may look less so.

The overarching theme of this work of mine is that we need to democratize and humanize finance in light of research on human behavior and the functioning of markets (Shiller 2011). Democratizing finance means making financial institutions work better for real people, dealing with the risks that are most important to them individually, and providing opportunities for inspiration and personal development. Humanizing finance means making financial institutions interact well with actual human behavior, taking account of how people really think and act.

Lionel Robbins, with his 1932 book An Essay on the Nature and Significance of Economic Science, has had the honor of inventing the most common definition today of economic science, of the unifying core idea that defines this science. He wrote then:

\[ \text{The economist studies the disposal of scarce means. He is interested in the way different degrees of scarcity of different goods give rise to different ratios of valuation of them, and he is interested in the way in which} \]

---

28Moss (2002) documents much discussion and experimentation with liability rules in the early nineteenth century, as with “double liability” that limits shareholders’ liability to twice their initial investment, or liability that ends when the shares are sold.
changes in conditions of scarcity, whether coming from changes in ends or changes in means—from the demand side or the supply side—affect these ratios.\(^{29}\)

The importance of prices in allocating scarce resources is an idea whose beginnings go back at least to Adam Smith in the eighteenth century, with his “invisible hand,” and there was a certain wisdom in Robbins’s framing of the entire field of economics around this idea. This wisdom still today is not fully apparent to the untrained public. Most people do not appreciate that all of our economic activities and all of our pleasures and satisfactions, and those of subsequent generations, are ultimately guided by prices of scarce resources as formed in the markets.

There is a problem, however, with the interpretation of economics that Robbins so persuasively gave. For his definition appeared to cast the economic problem exclusively as about scarcity of production resources, like energy and food, rather than also about scarcity of human intellectual and psychological resources. He casts the problem as man against nature, when in fact much of the economic problem is dealing with man against himself.\(^{30}\)

Long-term asset prices as they are observed today, prices of stocks, bonds, real estate, and commodities, and prices of derivative products such as futures, swaps, and options, and of other institutions like long-term insurance, are especially significant for economics, and especially problematic, since the scarcity that these prices represent is one that is never really objective and directly revealed today. Their levels are influenced by expectations of the distant, and generally nebulous, future. The market prices of speculative assets at any given time reflect, as is commonly asserted, both tastes and technology of that time. But they also reflect expected tastes and technology of the future, the likelihood of discovery of new sources of resources, or the technology to develop them. They also reflect sociology and social psychology, and anticipated future changes in these, in government policy such as taxation, and in other primary forces, such as changes in the inequality of incomes and likely social and governmental reactions to these, and the potential threat of wars and other catastrophes, and the likely use of, and policy toward, the assets in such times.

Fischer Black, in his 1984 presidential address before the American Finance Association, offered a new definition of market efficiency: he redefined an “efficient market” as “one in which price is within a factor of 2 of value, i. e., the price is more than half of value and less than twice value. . . . By this definition, I think almost all markets are efficient almost all of the time.”\(^{31}\)

And yet, even assuming he is somehow right, the existing efficient markets theory remains the fundamental framework from which many economic policy decisions, and decisions to innovate or not, are made. No one would seriously propose the elimination of stock markets, even if we all accepted Fischer Black’s impression as fact. So, why should we not consider other risk markets, markets that have not come into being yet just through accidents of history and timing of associated technological breakthroughs?

\(^{29}\)Robbins (1932, p. 15).

\(^{30}\)See, for example, Mullainathan and Shafir (2013).

\(^{31}\)Black (1986, p. 533).
Institutions can be redesigned so that they reframe people’s thinking, to the longer term and to things that are better subjects for their attention, by making markets for risks that are better tied to fundamentals people should be thinking about. Institutions that change framing might sometimes qualify as institutions providing a “nudge” as Thaler and Sunstein (2009) have put it, suggesting the right direction for people without being coercive. They base their thinking on a philosophy they call “libertarian paternalism” emphasizing the government’s providing incentives for appropriate behavior without coercion. Though our groundings in behavioral economics are similar, I wouldn’t stress that term, perhaps because it seems to suggest a top-down structure for society, with government at the top. The development of financial capitalism seems to be, or can be, a matter of the voluntary organization of most of society, integrating the activities of people in all walks of life in fulfillment of their diverse purposes. A vision for a better financial capitalism should not be top-down at all.

Some recent examples of financial innovation, examples of new experiments, can help clarify how innovation might help in an imperfect financial world. Consider first the social policy bonds proposed by Ronnie Horesh (2000) which have recently taken actual form by the social impact bonds first issued with the help of the nonprofit Social Finance, Ltd., in 2010 in the United Kingdom. These redirect speculative impulses into solving social problems over a meaningful horizon that is chosen by the issuer to be neither too short nor too long to allow effective solutions.

Consider also that new crowdfunding initiatives to create websites that allow large numbers of dispersed people each to share information and each to invest a small amount of money directly into new enterprises, without the usual financial intermediaries, have sprouted in many places around the world, with websites like kiva.org or kickstarter.com. They are poised after the US Jumpstart Our Business Startups Act of 2012 to transform venture capital. Such innovations can and certainly will cause some runaway bubbles and abuse of ignorant investors. But they could, on the other hand, if designed and regulated right, create a new way of arousing animal spirits and focusing informed attention onto venture investments. Crowdfunding may be more effective in funding ideas that are hard to prove, whose payoff is not immediate, that have a subtle social, environmental, or inspirational purpose beyond mere profits, and that only a small percentage of the population is equipped to understand.

Consider also the new benefit corporations that are now offered in 20 US states. They are amalgams of for profit and nonprofit corporations, fundamentally changing the mental framing that investors are likely to have of their investments in them, and encouraging both investors’ excitement and more idealistic thinking about these investments. The participation nonprofit business form that I advocated (2012), which makes nonprofits psychologically more similar to equity-financed business, would, if it is ever implemented, increase philanthropy and make it more effective.

These are only the beginning of the financial innovations that we might expect to see in our future, helped along by our improved understanding of behavioral finance, of mathematical economics, and steadily improving information technology. In

---

32 See http://benefitcorp.net/.
particular, it would seem that great benefit can be derived by expanding the scope of our financial markets, to allow trading of risks that really matter.

We might further democratize finance from the expansion of trading to include trading of other indices that have only recently come to be measured but that reflect real and important risks that matter to people. I have already alluded to the futures market for single family homes that was started at the Chicago Mercantile Exchange in 2006, and if that market becomes more successful it will eventually provide price discovery for a value of great personal importance to individuals, and might lead to a cash market for real estate that is not so woefully inefficient. The home futures market, if it became more successful, would facilitate the creation of many more risk management products, such as home equity insurance (Shiller and Weiss 1999b) or mortgages with preplanned workouts (Shiller 2012, 2014; Shiller et al. 2013).

Had there been a well-developed real estate market before the financial crisis of 2008, it would plausibly have reduced the severity of the crisis, because it would have allowed, even encouraged, people to hedge their real estate risks. The severity of that crisis was substantially due to the leveraged undiversified positions people were taking in the housing market, causing over 15 million US households to become underwater on their mortgages, and thus reducing their spending. There is no contradiction at all in saying that there are bubbles in the housing market and yet saying that we ought to create better and more liquid markets for housing.

Even further, I along with others have argued that a market for claims on the flow of gross domestic product, or other large macroeconomic aggregates, should be developed to help countries share their risks (Shiller 1993, 2003b, 2008; Athanasoulis and Shiller 2000, 2001; Kamstra and Shiller 2009); or markets for other significant economic variables like occupational incomes to share their livelihood risks (Shiller 1993; Shiller and Schneider 1998; Shiller 2003b).

Had the government debts of European countries taken the form of GDP shares, then most likely we would not have had the severe European sovereign debt crisis that started in 2009, for the countries would not have as big a short-run refinancing problem and would find their government obligations cushioned by declining obligations due to declining GDP. Had people sought protection for their own welfare by hedging themselves in occupational income markets, many of them would have suffered less in this crisis.

Examples of innovations that might reframe into better and longer-term thinking about fundamentals include the “perpetual futures” that I have proposed (1993),33 or the application of the concept of index participations developed by the American Stock Exchange in 1989 to flow indexes34 or the long-term MacroShares my colleagues and I once had striven to launch based on various indices35 or the markets for individual future dividend dates on stock price indices that Brennan (1999)

---

33 These are defined in Shiller (1993) in terms of a daily settlement formula involving both the change in settle price and another index representing a cash flow.
35 In 2006 our firm MacroMarkets LLC launched paired long and short 20-year oil MacroShares on the American Stock Exchange, with ticker symbols UCR for Up-Crude and DCR for Down-Crude. The securities traded from November 2006 to June 2008, and at one point reached US$1.6 billion in total value, but were not ultimately a success.
argued might “focus investor attention on the fundamentals that determine the value of the index rather than simply on the future resale value of the index.”36

The development of inflation-indexed bonds, which have gradually grown in importance over the last half century worldwide, are an important past success, but as yet an incomplete one. Such markets, and other indexing institutions, might be enhanced by further deliberate changes in psychological framing. If inflation-indexed units of account, which create an easier way in our language to refer to indexed quantities, were created and widely used, they would help people around their money illusion which inhibits intelligent design of contracts around the real outcomes that really matter. I have been advocating the proliferation of these units of account where they first began in Chile (2002), in the United States (2003), and the United Kingdom (2009).37 Their widespread use might have helped prevent the real estate bubble that preceded the current financial woes, a bubble that was likely helped along by the widely held impression that single family homes have historically shown high real capital gains when in fact over the last century the gains overall have been only nominal and hence illusory (Shiller 2005).

We want such innovations, if not exactly the ones I and others have been advocating to date, because their predecessor innovations, the financial institutions we already have today, have brought such prosperity, despite the occasional big disruptions caused by bubbles and financial crises. There is no economic system other than financial capitalism that has brought the level of prosperity that we see in much of the world today, and there is every reason to believe that further expansion of this system will yield even more prosperity.

The patterns of behavior that have been observed in speculative asset prices are consistent with a view of market efficiency as a half-truth today and at the same time with a view that there are behavioral complexities in these markets that need to be met with properly engineered financial innovations and financial regulations.

Changes in our financial institutions that take the form of creative reinventions in the kinds of risks traded, that change the psychological framing of the things traded, and that change our social relations with business partners and adversaries, can make financial markets less vulnerable to excesses and crashes and more effective in helping us achieve our ultimate goals.

REFERENCES


36 Brennan (1999, p. 12). Since 2008, dividend futures markets for stock price indices have appeared on a number of European and Asian exchanges, though it is not clear that these new markets have had much of the desired effect of reframing investors’ thinking.

37 Chile created its Unidad de Fomento (UF) in 1967, still in use there today, http://valoruf.cl/. This innovation helps deal with public resistance to indexation (Shiller 1997).


Morris, Stephen. 1996. “Speculative Investor Behavior and Learning.” Quarter-


