MEMORANDUM

FROM: César Orosco
DATE: September 18, 2017
RE: EQUITY RISK PREMIUM

No sooner had Larry Siegel’s latest landed on my desk than *The Economist* arrived.

---

**Buttonwood | Getting the long term right**

**The trickiest part of forecasting the distant future concerns valuation**

What is the right way to invest for the long term? Too many people rely on past performance, picking fund managers with a “hot” reputation or backing those asset classes that have recently done well. Just as fund managers cannot be relied on to be consistent, returns from asset classes are highly variable. The higher the initial valuation of the asset, the lower the future returns are likely to be.

That is pretty clear with government bonds. Anyone buying a bond with a yield of 2% and holding it until maturity can expect, at best, that level of return (before inflation) and no more. (There is a small chance the government might default.) With equities, the calculations are not quite so hard-and-fast. Nevertheless, it is a good rule-of-thumb that buying shares with a low dividend yield, or on a high multiple of profits, is likely to lead to lower-than-normal returns.

So a sensible approach to long-term investing would assess the potential returns from asset classes, given their valuations and the fundamentals, and allocate assets accordingly. That is what GMO, a fund-management company, has been trying to do for decades. It has made some common-sense assumptions about the fundamental drivers of returns and then assumed that valuations would return to average levels over a seven-year period.

In one sense, this process has been a success. The assets that GMO thought would perform well have offered relatively high returns, the assets it thought would perform badly have offered lower ones (see top chart). But if the ranking has been correct, the level of return has not been. Assets that GMO thought would yield a negative return of 10% to 5%, for instance, have in fact suffered average losses of only 2.8%.

GMO’s forecasts have been pretty accurate for asset classes such as emerging-market bonds and international (non-American) shares; annual returns have been within 1½ percentage points of its forecasts. But for American equities, GMO was too gloomy, underestimating returns by around four percentage points a year.

The reason for this error is pretty clear. Equity valuations have not returned to the mean, as GMO thought they would, but have stayed consistently above their historical levels. GMO was fairly accurate in its forecast for dividend growth, but its erroneous estimation of valuation accounted for all the forecast error.

There are two possible conclusions. One is that GMO is simply wrong about mean reversion. Equities have moved to a new, higher valuation level. This sounds uncomfortably like the famous quote from Irving Fisher, an economist, before the 1929 crash, that stocks had reached a “permanently high plateau”. But there is some justification for a valuation shift: American profits have been high, relative to GDP, for a long period of time. This may be a result of monopoly power in some industries (see Free exchange), or perhaps of the reduced bargaining power of workers in an age of globalisation.

A more obvious argument is that, with yields on cash and government bonds so low, investors are willing to pay a high price for equities because they represent their only hope for decent returns. But given the low level of dividend yields and the sluggish rate of economic growth, profits will have to keep rising as a proportion of GDP to allow high equity returns to continue. That seems unlikely. Either there will be a political reaction—governments will clamp down on firms in response to public unrest—or, more prosaically, tighter labour markets mean that wage growth will start to erode profits.

Either way, it is understandable that GMO does not want to give up on the idea of mean reversion just yet. Its latest forecasts are pretty downbeat (see bottom chart). The real returns from most asset classes are expected to be negative; only emerging-market equities offer a decent return. Investors who disbelieve those forecasts are in essence betting that things will be “different this time”. That is certainly possible but it requires a lot of faith. 

---

The Economist blogs/buttonwood
Few can distill ideas to their absolute essence and organize competing points of view in a cohesive framework the way Larry does. For anyone interested in capital market returns, his review is required reading (and will save you even more reading)!

CAO
orosco@ajopartners.com

gce

P.S. Leave it to Larry to make the often overlooked Bibliography section also a must read. Journey with him to the end and you will see.
The equity risk premium (ERP), or equity premium, is the difference in expected or realized return between an equity index and a reference asset, where the latter is usually a bond or bill portfolio considered to be “riskless.” In the modern literature and in investment management practice, ERP usually means “expected ERP” and we will stick to that convention, reserving the phrase “realized ERP” for any backward-looking or historical measure.

The ERP is widely acknowledged as the most important variable in finance. It is useful for:

- Knowing what returns to expect from each major asset class and from portfolios of securities or asset classes
- Lifecycle and retirement planning (so one can estimate how much to save and invest in the hope of achieving a given standard of living in retirement), and
- As a component of the opportunity cost of capital or required rate of return in corporate finance

An estimate of the ERP is required for essentially all asset allocation models and is central to the practice of investment management and asset-liability management. ERP estimates thus strongly affect the asset allocation decisions of individual investors and institutional investors, including pensions, endowment funds, foundations, and insurance companies.

---

1 This article will be published in the fall of 2017 by the CFA Institute Research Foundation as part of its Literature Review series, accessible at http://www.cfapubs.org/loi/rflr. It should be treated as copyrighted material.

Laurence B. Siegel is the Gary P. Brinson director of research at the CFA Institute Research Foundation, Charlottesville, VA, lbsiegel@uchicago.edu. The author thanks P. Brett Hammond, research leader at Capital Group (Los Angeles), for his top-level editorial assistance and extensive suggestions. Zhiyi Song, CFA, PhD, allowed me to recycle some of the ideas and language in the predecessor article to this one, “The Equity Risk Premium: An Annotated Bibliography” (CFA Institute Research Foundation, 2007); the section on the equity premium puzzle is almost entirely his (although I have shortened it), as are many of the annotations. I also thank various anonymous interviewees, including some whose work is cited herein.

2 Occasionally the reference asset is “inflation,” that is, a hypothetical asset returning the rate of consumer price inflation as measured by some index.

3 We would argue that no asset is completely riskless.
APPROACHES TO ESTIMATING THE ERP

This article is organized by theme, roughly in the historical order in which the themes first appeared in the literature. There are three broad categories of approaches to estimating the ERP:

1. Methods based on a dividend discount model (DDM), earnings discount model, or cash flow-to-the-investor discount model: forward-looking methods with their roots in discounted cash flow (DCF) analysis wherein the value of an asset is regarded as the present value of the cash flows the asset is expected to generate.

2. Methods based on extrapolating past trends, in particular the spread between realized stock and bond or cash returns, into the future: retrospective methods.

3. Methods based on a macroeconomic model of the way that investors require compensation for risk.

In past literature these have been called, respectively, supply, equilibrium, and demand models. The DDM is a supply model because it focuses on ways that companies generate cash with which to reward investors. The macroeconomic model is a demand model because it asks what excess return investors need to induce them to take equity risk. The retrospective method can be regarded as an equilibrium model because it relies on prices at which the market actually traded, reflecting the intersection of supply and demand curves.

EARLIEST ESTIMATES

The earliest estimates of the ERP are derived by estimating the expected return on an equity portfolio using dividend discount model (DDM), then subtracting the expected return or yield on the riskless asset. This “DDM” approach, which made a comeback at the end of the twentieth century, is the most widely used one now.

FUTURE EQUALS PAST

The next step was taken by researchers who measured the realized ERP, asserting that the realized ERP was the best estimate of the expected ERP. In their view, neither the amount of risk in the market or the “price of risk” (the return investors require and expect to receive for taking a given amount of risk) changes much over time; that is, the return-generating process for equities (in excess of the riskless rate) is stable or stationary. We call this method the future-equals-past approach.

The future-equals-past approach suffers from the following flaw: the higher the market rises, the higher the estimate of future returns given by the method. This is contrary to intuition, where you expect a low return (on any asset) if you pay a high going-in price.

---

for it. Consider, for example, a bond: if the past return is 10% per year because interest rates have been falling, say, from 5% at the beginning of the holding period to 1% at the end, is the expected return 10% or 1%? It is the latter.

In addition, future-equals-past assumes that markets are fairly priced and does not allow for investigation into the possibility that they are not. This possibility became a primary focus of research once the future-equals-past method lost its preeminence.

**THE MACRO APPROACH OR “EQUITY PREMIUM PUZZLE”**

Academics starting around 1985 began to question why the realized ERP, and apparently also the expected ERP, was so large when certain aspects of macroeconomic theory suggested it should be much smaller — that is, other tradeoffs between risk and reward in the economy implied that investors did not require nearly as large a risk premium as they had been getting.

This “equity premium puzzle” literature, while extensive and contentious, turned out to be something of a dead end because the ERP, while arguably smaller than it once was, is still much larger than the puzzle literature says it should be. We nevertheless take this literature seriously and document it in our “Puzzle” section.

**THE DDM COUNTERREVOLUTION**

A substantial innovation occurred in the 1980s when several researchers found the ERP to be time-varying. This literature spawned a mountain of research on the time-series behavior of equity market valuation measures, particularly price-earnings (P/E) ratios.

The P/E-related research asks, among other questions, what the best definition of “earnings” is for forecasting future returns. The CAPE (cyclically-adjusted price-earnings ratio), which smooths earnings data by averaging them over long periods, typically ten years, has become the most popular measure. (The P/E and CAPE are relevant to ERP estimation because, if the ERP is time-varying, these statistics provide a way to get continuously updated measures of the expected return on equities; one can then subtract bond or bill yields to arrive at the ERP itself.) This thread, which we call “time-varying premia,” continues today as the predominant one in ERP research.

A branch in the time-varying ERP tree asks what, besides earnings, might accurately measure the desirability of an equity investment. The most important alternative is payout, or “cash flow to the investor” — that is, dividends plus other cash flows such as those from share buybacks.

**OTHER WORKS**

While most of the work that has been done on the ERP relates to the United States, the underlying issues are the same everywhere. We review literature that extends this work to international markets.

Finally, we list and comment briefly on other literature reviews, compilations, and aggregative works.
**First Stirrings**

Edgar Smith [1924] seemed to intuit the equity risk premium. He presented evidence that stocks had high returns, realized or expected (he did not make the distinction), relative to other, primarily fixed-income assets. The Harvard professor John Burr Williams [1938] was the first to state that the value of a firm is the discounted present value of all of its future dividends. He wrote, “Earnings are only a means to an end [dividends], and the means should not be mistaken for the end.”

Williams’ discounted cash flow (DCF) formula, familiar to all business students, represents the origin of risk-premium thinking because the discount rate, in order to be useful for valuing stocks, must be a risky discount rate that is higher than the riskless rate by an amount (the equity risk premium) that compensates the investor fairly, but not more than fairly, for the risk of the stock.

Myron Gordon and Eli Shapiro [1956], building on Williams’ work, formalized the notion of a risky discount rate and equated the expected return on an equity with the “required rate of profit.” This principle is still the foundation of corporate finance, which asserts that the market for an asset (say, an equity) is in equilibrium when the expected return on the asset equals the required return, that is, the return that investors demand as fair compensation for the asset’s risk.

**Future equals past**

But these early works did not lead directly to estimates of the ERP that were usable for asset allocation, capital budgeting, and other uses to which the premium is now put. Ibbotson and Sinquefield [1976a] made explicit estimates of the premium by calculating, for as far back in history as high-quality data allowed, the difference between the realized total returns on an equity index and the realized total returns on a bond or bill (cash) portfolio. The logic was that, over time, investors “conform their expectations to that which proves to be realizable” in fact, so that the historical return (in excess of the riskless rate) is a fair or equilibrium estimate of the return (in excess of the riskless rate) that investors should expect going forward.

Ibbotson and Sinquefield, then, decomposed historical returns on an equity index into a part attributable to the riskless rate and a part attributable to the equity premium. The arithmetic mean of the equity premium part is assumed to be stationary — that is, the same in the future as in the past. Thus, if equities had beaten riskless Treasury bills by an arithmetic mean margin of 7 percent a year over the historical measurement period, which was usually 1926 through the then-current time, then equities were forecast to beat bills by the same amount in the future.

---


The arithmetic mean expected total return on equities was then calculated as the sum of the forward-looking riskless rate (that is, the yield on riskless bills or bonds) and the arithmetic mean expected ERP.

Reflecting on Ibbotson and Sinquefield’s pioneering work, I wrote,

Hadn’t anyone before... Ibbotson and... Sinquefield estimated the equity risk premium? Of course the thought had occurred to many, but the preexisting methodology — to use a kind of Dividend Discount Model (DDM) for the aggregate of all stocks in the market — gave forecasts, or estimates of the ex ante or expected risk premium, not backward looks at history.

Hindsight showed that DDM-based forecasts had been much too low. A typical DDM estimate of the forward-looking, or expected, equity risk premium over bonds was in the range of 2 to 3 percent. In contrast, Ibbotson [and Sinquefield] showed that stocks had out-returned intermediate-term Treasury bonds by much more, 5.4 percent, using 1926 to 1979 as the measurement period.8

Ibbotson and Sinquefield’s work was tremendously influential, led to the establishment of a firm (Ibbotson Associates) that would later be acquired by Morningstar, and was updated by Morningstar in yearbook form until 2015 and thereafter by Duff & Phelps [Duff & Phelps 2017].9 Their method is still the way that many finance professors, investment management and sales executives, and others make their long-run forecasts. However, over roughly the last quarter-century, other methods — principally based on a forward-looking discounted cash flow (DCF) model, such as the dividend discount model (DDM) — have become competitive and even dominant.

THE DDM COUNTERREVOLUTION BEGINS

As just noted, the future-equals-past method was the principal way of estimating the ERP for a long time after Ibbotson and Sinquefield’s early studies. However, a 1984 paper, which was mostly ignored at the time but which would later become influential, called into question the relevance of its forecasts. Jeffrey Diermeier, who would later serve as president of the CFA Institute, wrote the paper with Roger Ibbotson and myself.

We argued that (1) corporate earnings could not indefinitely grow faster than the overall economy, or else there would eventually be nothing left for labor, government, and other claimants; and (2) the PE ratio could not rise indefinitely either. As a result, the growth rate of the economy — that is, of GDP — is the hypothetical upper limit of the very-long-term rate of price return on equities. In addition to that, the investor receives dividends. See Diermeier, Ibbotson, and Siegel [1984]).

---


9 Starting about 2015, Morningstar discontinued “future equals past” estimates of the ERP in its updates of the Ibbotson yearbook, noting that DDM-type forecasts are more accurate and more theoretically justifiable.
This argument asserts that a DDM is the right way to think about the ERP. While this idea remained dormant for some time, it would constitute the main thrust of ERP estimation in the 1990s and thereafter.

**THE EQUITY PREMIUM “PUZZLE”**

In the 1980s, while practitioners were debating whether the ERP was low (3% or 4%, as suggested by DDM methods) or high (more than 5%, as obtained by extrapolating historical data), a group of academics was wondering why the ERP was not trivially more than zero. Mehra [1983], and Mehra and Prescott [1985], describe a “puzzle” in which the historical ERP during the period of 1889-1978 (or any other similarly long period, such as 1926 to the present) was too high, by at least an order of magnitude, to be explained by standard “general equilibrium” or “macroeconomic” asset-pricing models.

Using these models, such a high premium can only be explained by a very high coefficient of risk aversion, one in the range of 30 to 40. (The risk aversion parameter describes a given individual’s tradeoff between the amount of risk taken and the amount of additional return he or she requires as compensation for taking that risk.) Risk aversion parameters observed in other aspects of financial behavior are around 1. So, Mehra and Prescott argued, either the model used to describe investors’ behavior is flawed or equity investors have received a much higher return than they expected.

We call the asset-pricing models referenced by Mehra and Prescott [1985] “macroeconomic” because they originated in that specialty, but more importantly to distinguish them from asset-pricing models commonly used in investment finance — such as the Capital Asset Pricing Model, the three-factor Fama-French model, and arbitrage pricing theory — that are silent on the absolute size of the risk premium (in fact, requiring it as an input) and that distinguish instead among the expected relative returns on specific securities or portfolios.

**RESOLVING THE PUZZLE**

Research on the question of why the realized equity premium was so large can be grouped into three broad categories: (1) studies alleging bias in the historical data, (2) studies suggesting improvements in the macroeconomic model, and (3) studies that raise behavioral finance, lifecycle, and other issues.

**BIASES IN HISTORICAL DATA.** Potential biases in the historical data include survivorship bias, transaction and tax costs, and the mixing of expected and unexpected components of past returns.

*Survivorship bias.* Brown, Goetzmann, and Ross [1995] argued that the historical equity premium calculated using U.S. data is likely to overstate the true (expected) premium because the U.S. stock market turned out to be the most successful in world history. However, Dimson, Marsh, and Staunton [2008] examined stock and bond returns using data from 1900 to 2005 for 17 countries and concluded that the high historical equity premium obtained for the United States is comparable with that of other countries.

*Transaction costs, regulations, and taxes.* McGrattan and Prescott [2001] suggested that the historical equity premium was higher than expected mainly because of a large
run-up in the equity price caused by the sharp decline in the tax rate on dividends. In their 2003 article, they claimed that the equity premium is less than 1 percent after accounting for taxes, regulations, and costs. To this thought, we’d add that index funds were not available to investors over the long periods studied by historical researchers; thus equity investors earned returns lower than those of the index by the amount of (1) the explicit transaction and holding costs involved in forming portfolios; and (2) the implicit cost of not being diversified.

Unanticipated repricing of equities. Bernstein [1997] suggested that, because equities started the sample period (which begins in 1926) at a price-to-earnings ratio (PE) of about 10, and ended the period at a PE of about 20, the actual return on equities was higher than investors expected or required. Thus, the historical return overstates the future expected return. This finding was bolstered by Fama and French [2002], who used the DDM to show that investors expected an equity risk premium of about 3 percent, on average, from 1926 to the present.

IMPROVEMENTS IN THE THEORETICAL MODEL. The second broad category of research on the equity risk premium is a large body of literature exploring a variety of improvements in the original Mehra and Prescott [1985] model.

Rare events. Rietz [1988] suggested that the ERP puzzle can be solved by incorporating a very small probability of a very large drop in consumption. If such a probability exists, the predicted equity premium is large (to compensate investors for the small risk of a very bad outcome). Mehra and Prescott [1988] countered that Rietz’s model requires a 1 in 100 chance of a 25 percent decline in consumption to reconcile the equity premium with a risk aversion parameter of 10, which is less than the risk aversion required to explain the realized premium — and that the largest aggregate consumption decline in the last 100 years was only 8.8 percent.10

We’d remind these debaters that, according to Cooper and John [2012], in the United States “from 1929 to 1933, real GDP decreased by 26.5 percent, while consumption decreased by 18.2 percent” (§22.4).11 Mehra and Prescott’s 8.8% was the consumption decline in just one year of a multi-year decline.

Campbell, Lo, and MacKinlay [1997] pointed out that “the difficulty with Rietz’s argument is that it requires not only an economic catastrophe, but one which affects stock market investors more seriously than investors in short-term debt instruments” (p. 311).12 Barro [2006] extended Rietz’s model and argued that it does provide a plausible resolution of the equity premium “puzzle.”

Borrowing constraints and life-cycle issues. Constantinides, Donaldson, and Mehra [2002] introduced lifecycle and borrowing constraints. They argued that, as the correlation of equities with personal income changes over the life of the investor, so too does the attractiveness of equities to that investor. The young, who should borrow to smooth consumption and to invest in equities, cannot do so. Therefore, equities are priced almost exclusively by middle-aged investors, who find — or at one time found — them to be unattractive. Thus, equities are underpriced and bonds are overpriced, producing a higher ERP than the puzzle literature predicts.

Behavioral concerns. A large swath of behavioral finance literature argues that the combination of “myopic” loss aversion and narrow framing can help to resolve the equity premium puzzle, including works by Benartzi and Thaler [1995], Barberis, Huang, and Santos [2001], and Barberis and Huang [2006].

TIME-VARYING PREMIA AND THE DDM COUNTERREVOLUTION

In one of the sharpest academic-practitioner divides in memory, some academics still consider the puzzle literature relevant while almost no practitioners do. In addition, the future-equals-past method is rarely used by sophisticated practitioners, showing up mostly in the marketing literature of private wealth advisors who are trying to sell equities. So the DDM-based approach is the only one with any real traction after the turn of the millennium.

While some practitioners had long used DDM-type estimates of the ERP, Campbell and Shiller, in the late 1980s, were really the first to re-establish the DDM as a respectable challenger to the then-dominant future-equals-past method. Their work spawned a vast literature that is exclusively forward-looking — that is, it focuses on the expected rather than the realized ERP. This literature asserts that, like most DDM estimates, the ERP is time-varying and countercyclical: the ERP is high when the market is low and vice versa.

As noted earlier, the future-equals-past method, in contrast, is procyclical — it paradoxically gives higher forecasts after each market move upward and lower forecasts after each move downward. (An interesting contrast of investors’ procyclical views with the DDM’s countercyclical forecasts is in Greenwood and Shleifer [2014]).

This procyclicality proved to be the method’s undoing. As of 1999, it was forecasting a greater than 12% annual return, an absurdity given the already bubble-like level of the market. So, around that time, the popularity of the future-equals-past method waned and acceptance of the DDM and allied approaches grew. Because the DDM had also been the preeminent method before Ibbotson and Sinquefield, we refer to this shift in thinking as the DDM counterrevolution.13

Valuation levels and subsequent stock returns

Campbell and Shiller [1988] “found that valuation ratios are positively correlated with subsequent returns and that the implied predictability of returns is substantial at longer horizons” (Campbell [2007], p. 1). So much for perfectly efficient markets! If returns can be predicted from valuation levels then return expectations are not, or should not be, constant; thus (holding the riskless rate constant) the ERP is not constant either. There is information in valuation levels, then, potentially useful for timing the market and almost certainly useful for making periodic adjustments to the ERP assumption used in asset allocation and long-range planning.

Around that time, Fama and French [1988] came to a similar conclusion. They found that dividend yields were positively related to expected stock returns. This is the same as saying that high valuations (low dividend yields, that is, high price-to-dividend ratios) portend low stock returns and vice versa.

Once Jeremy Siegel [1994] and Peter Bernstein [1997], both best-selling authors with strong academic credentials, jumped decisively on the DDM bandwagon (see our discussion of Bernstein’s work earlier), other works pursuing the same theme came in a flood. They include Campbell and Shiller [1998], Arnott and Bernstein [2002], Shiller [2000], Asness [2000, 2003], and Fama and French [2002] (mentioned earlier in the Puzzle discussion). As the field matured, other, more integrative works were produced, including Cochrane [2011] and Ilmanen [2011].

Two influential books

Among practitioners, the most influential of these works were Siegel’s and Shiller’s books, respectively titled Stocks for the Long Run and Irrational Exuberance. Sometimes portrayed as rivals, the two authors are actually close personal friends who have vacationed together with their families and who enjoy debating the fine points of their views on markets.

CAPE method. Shiller’s book, in particular, has spawned a literature on the valuation method it espouses, called CAPE (cyclically-adjusted price-earnings ratio). The CAPE literature is relevant to ERP estimation because CAPE is just an “improved” PE ratio — which, under carefully constrained conditions, is the inverse of the real expected return on a stock or stock portfolio.

Thus, if the CAPE or PE ratio of a portfolio (say, an index) is 25, the real expected return is $1/25 = 4\%$, and one can then subtract the real riskless rate (say, 1\%, which is roughly the rate as of this writing) to arrive at the ERP (in this example, 3\%). Jeremy Siegel [2016] sets forth a constructive critique of the CAPE method, saying we should emphasize more recent data, rather than the entire history, because accounting for the goodwill component of corporate earnings became more conservative around 1990. Adjusting for the accounting change raises the equity premium forecast.

Market timing. As suggested above, if the ERP is time-varying, then of course one could use that information to time the market. For obvious reasons, the literature on market timing intersects with the literature on estimating the value of the ERP at a given time. Ilmanen [2016] focuses on the time-varying aspect of the ERP and other risk premia, but because his discussion also deals with long-term expectations, it is
included here. Market timing per se is outside the scope of this review, but practitioners widely use methods such as those described by Ilmanen to manage assets.

**Cash Flow to the Investor**

The “payout,” or “cash flow to the investor,” literature relies on Miller and Modigliani [1961], whose work implies that, in the words of Straehl and Ibbotson [2017], “investors should be indifferent about whether they receive distributions via dividends or buybacks as well as how they participate in a buyback — that is, by receiving cash from tendering their shares or by receiving an increased proportion in the company” (p. 2).14

If this is the case, then explicit (cash) dividends are irrelevant and only total cash payout to the investor, including buybacks as well as dividends, is relevant for equity valuation. Diermeier, Ibbotson, and Siegel [1984] rely on this principle, as do Grinold and Kroner [2002] and Grinold, Kroner, and Siegel [2011] (which was discussed earlier).

The Grinold et al. articles adjust dividends for “net new issues,” that is, the number of shares issued by companies in secondary public offerings minus the number of shares retired through buybacks and other corporate actions. This method brings together (1) the payout literature and (2) the dilution analysis performed by Bernstein and Arnett [2002], wherein the authors find that, in order to achieve the earnings growth that has been observed, shareholders have had to suffer dilution amounting to a large 2% per year — “dilution” referring to a decrease in the share of a company represented by a given number of shares. This dilution, if continued in the future, reduces the ERP.

But Straehl and Ibbotson [2017] were the first to really complete the payout analysis. They show that total payouts — in their formulation, dividends plus buybacks, not dividends alone — explain long-run stock market returns.15 They propose a new valuation measure, CATY (cyclically adjusted total yield), analogous to CAPE but constructed from “total yield” (payouts) rather than earnings, that “predicts changes in expected returns at least as well as the [CAPE].”16 (The analysis is still not quite complete because total yield should include cash takeovers as well as dividends and buybacks, but the authors did not have data for cash takeovers.)

As with the CAPE literature, the payout or CATY literature ties back to the ERP through the fact that the ERP can be calculated simply by subtracting the bond or bill yield from whatever expected total return on the stock market is implied by the CAPE or CATY analysis. Estimating the ERP and estimating the expected stock market return are essentially the same problem, because the two estimates differ by an observable constant (the riskless rate of return).

---


15 In an article in progress, I argue that cash takeovers are a form of buyback and should be added to the total-payout calculation.

16 From the authors’ abstract.
Yet the payout literature is contentious because it relies on a satisfactory disentangling of earnings, earnings per share, number of shares, new issues, dilution, dividends, retained earnings, buybacks, and takeovers. These concepts, governed by accounting identities, seem easy until one tries to interpret them for estimating expected returns and the ERP. Then they become difficult. While analysts perform this intricate analysis for individual companies with DCF models (by constructing measures such as EBITDA), it may be daunting to do this in the aggregate. In an inversion of the classic framework where dividends are easy to forecast and capital gains hard, the payout literature shows that even the income part of the return, of which dividends are a key element, is subject to interpretation and controversy.

Why the increased interest in payout? Brav, Graham, Harvey, and Michaely [2005] report that a 1982 change in Securities and Exchange Commission (SEC) rules reduced the legal risk of repurchases. Since that time, dividend yields have fallen and buybacks have soared. In particular, according to Cornell, Arnott, and Moroz [2005], “managers behave as if there is a significant capital market penalty associated with cutting dividends, but not with reducing repurchases. Accordingly, dividends are set conservatively and repurchases are used to absorb variation in total payout.” The resulting increase in buybacks makes it important to measure them as part of total payout rather than relying, as analysts in the last century generally did, on dividends.

OTHER METHODS

Our review of ERP estimation methods is not exhaustive. Duarte and Rosa [2015], making one-year rather than long-term forecasts, catalogue 20 models and “find that an optimal weighted average of all models places the one-year-ahead ERP in June 2012 at 12.2 percent, close to levels reached in the mid- and late 1970s, when the ERP was highest in the study sample.” We now know that this forecast was roughly correct over the subsequent five years, but it is way too high as a long-term expectation. This result suggests that the method might be used for making the medium-term forecasts needed for dynamic or tactical asset allocation (timing) decisions.

CREDIT MARKETS AND THE ERP

An alternative approach to estimating the ERP is to look at credit markets. Equities per se don’t have observable expected returns, but equity-like risky bonds do; the expected return is the yield, minus an allowance for defaults. (The default allowance must necessarily be an estimate or forecast.) By extrapolating the risk-return relationship for credit bonds up to the risk or beta of equities, one can get a usable ERP number.

The literature on this question is well represented by Berg and Kaserer [2005], who use credit default swap (CDS) spreads instead of bond yields because of their greater accuracy. Their results for the U.S. ERP range from 5.16% in 2004 to 7.18% in 2005; they note that, while the forecasts are high, these are upper limits, not midpoint estimates.

---

INTERNATIONAL ISSUES
The first efforts at measuring long-run equity returns in global markets were Ibbotson, Siegel, and Love [1985] and Brinson, Diermeier, and Schlarbaum [1986]. But neither of these explicitly estimated an ERP (although they made such estimation possible using a future-equals-past method). It took until the turn of the millennium for academics’ attention to turn to the global equity market and its risk premium in a meaningful way.

Jorion and Goetzmann [1999] tested the concept of survival bias, which asserts that ERP estimates taken from successful countries, such as the U.S. or U.K., are upwardly biased because one could not know at the beginning of the period studied which countries’ markets would survive and which would fail or almost fail due to war, nationalization, or for other reasons. This is a key issue in the estimation of any variable from observed historical data.

The authors “collect a database of capital appreciation indexes for 39 markets going back into the 1920s. Over 1921 to 1996, the U.S. had the highest real return of all countries, at 4.3%, versus a median of 0.8% for other countries. The high equity premium obtained for the U.S. therefore seems to be the exception rather than the rule.”

Dimson, Marsh, and Staunton [2002, 2017] have a slightly different take on survival bias. They document, for a large assortment of countries, the annual returns on equities, bonds, and bills, over a very long period: 1900 to the present. They also document exchange rates and inflation rates so that real returns can be compared across countries. Like Jorion and Goetzmann, they show that survival bias is a significant factor in interpreting historical equity returns: an index composed of just countries that survived the twentieth century, with its wars and nationalizations, outperformed an unbiased index composed of countries that had markets in 1900.

However, the U.S. — one of the highest-returning markets — outperformed other surviving markets by only a modest margin. Equities, representing aggressive bets on the future, had the best returns in every country, representing the “triumph of the optimists” over pessimists who sought, through fixed-income investing, to defend their wealth positions against unforeseen disasters. Thus survival bias is not as large a factor as one might naively guess.

Jeremy Siegel [1994, 2014] also weighs in on survival, noting that stocks beat bonds even in countries where markets were almost extinguished by war and inflation; in Germany and Japan, for example, stocks survived but bonds were ruined entirely.

LITERATURE REVIEWS, COMPILATIONS, AND OTHER AGGREGATIVE WORKS
CFA INSTITUTE EFFORTS
In 2002, the Association for Investment Management and Research (now renamed the CFA Institute) convened a group of academic and practitioner experts on the equity risk premium and published the ensuing discussion (Leibowitz et al. [2002]; see the Annotated Bibliography for the link). There is no printed version, so some extremely valuable articles are left out of this literature review. Estimates of the ERP spanned the range from 0.0% to 5.0%, excluding survey results which ranged as high as 7%. The average of the experts’ estimates was 3.7%.
Hammond, Leibowitz, and Siegel [2011] present a reconvening of the group in Leibowitz et al. [2002], this time by the CFA Institute Research Foundation, and with some additions and deletions of personnel. Several of the individual articles in it are referenced separately in the current essay. Remarkably, in the decade since the previous convocation, the experts’ ERP estimates converged tightly to 4%, plus or minus a small amount.

**ADDITIONAL CONTRIBUTIONS**

Additional elements of the ERP literature include Goetzmann and Ibbotson [2006], Campbell [2007], De Long and Magin [2009], Cochrane [2011], Damodaran [2016], and Song [2007]. Goetzmann and Ibbotson’ book, entitled *The Equity Risk Premium*, is an indispensable collection of the two Yale professors’ works, with many co-authors, over more than 40 years. Several of the articles are referenced separately in the current essay. Song [2007] is the predecessor to this article. It emphasizes the “puzzle” more than I have, and is a valuable reference for readers interested in covering that literature in greater detail.

**CONCLUSION**

It is important to study and estimate the equity risk premium (ERP) because it underpins some of the most important financial and investment decisions a person or organization can make. Because the ERP cannot be observed directly, it must be estimated using one of a number of indirect approaches or models.

ERP models have gone through a number of fashions, sometimes called *regimes*, since the idea of estimating the ERP first came to prominence almost a half-century ago. At first, estimates of the equity risk premium, arrived at casually, tended to be low. Then, in the 1970s, Ibbotson and Sinquefield launched a period in which the equity risk premium was expected to be high. This period lasted between a decade and a quarter-century, depending on one’s viewpoint about when the DDM counterrevolution became fully established. Since the counterrevolution, the DDM approach seems to have prevailed and low to moderate estimates of the ERP have predominated.

What will happen in the future? While no one knows for certain, a low-return environment, sustained for a long enough time, creates the conditions for a high-return environment. But that has not happened yet. Market prices and valuation ratios suggest that low to moderate expected equity risk premia will prevail for some time.
The expected equity return equals the dividend yield, plus dividend growth, plus the expected change in valuation, if any. As of year-end 1925, investors expected about 5.1 percent (about 1.4 percent more than the bond yield). The subsequent positive surprise was because of four historical accidents: (1) bonds had unanticipated losses; (2) valuations quadrupled, as measured by the price-to-dividend ratio (P/D); (3) the market survived; and (4) accelerated growth in real dividends and earnings occurred because of regulatory reform. These observations are used to construct a framework for estimating the equity risk premium at each point in time, including the present. The “normal” equity risk premium, or historical average of what investors were actually expecting, is 2.4 percent, and the current (2002) equity risk premium is around zero.

Changes in the expected equity risk premium are explained by changes in the relative volatilities of the two assets (stocks and bonds). The low ERP at the time of writing is consistent with low volatility in the stock market and high volatility in the bond market. The author writes, “This model explains 1871-1998 data extremely well.” Interestingly, this finding holds over periods (before modern times) when volatility was not widely measured or understood to be a factor in asset returns.

The “Fed model,” so called because the U.S. Federal Reserve has sometimes used it to assess market valuation levels, compares the earnings yield (E/P or reciprocal of the PE ratio) on a stock market index to the yield on Treasury bonds. Asness argues that the model is incorrect because it mixes real and nominal quantities. Corporate earnings are “real,” varying with inflation, and bonds are nominal. Thus, the proper comparison is between earnings yields and real interest rates.

The authors review the behavioral approach to understanding the ERP puzzle. The key elements of this approach are loss aversion and narrow framing, two well-known features of decision making under risk in experimental settings. By incorporating these features into traditional utility functions, Barberis and Huang show that a large equity premium and a low and stable risk-free rate can be generated simultaneously, even when consumption growth is smooth and only weakly correlated with the stock market.

This paper proposes a new approach for pricing assets by incorporating two psychological ideas into the traditional consumption-based model. Investors are assumed to be more sensitive to losses than to gains, and their risk aversion changes over time depending on their prior investment outcomes. The authors show that this framework can help explain the high historical equity risk premium.

This paper extends the analysis of Rietz [1988] and argues that it does provide a plausible resolution of the ERP puzzle. The author suggests that the rare-disasters framework (i.e., the allowance for low-probability disasters proposed by Rietz) can explain the ERP puzzle while “maintaining the tractable framework of a representative agent, time-additive and iso-elastic preferences, and complete markets” (p. 823). These technical terms refer to assumptions that are embedded in Mehra and Prescott [1985] and that are considered standard in general equilibrium or macroeconomic models.


This article proposes an explanation for the equity premium based on two concepts from the psychology of decision making. The first concept is called “loss aversion,” meaning that investors are more sensitive to losses than to gains. The second concept is called “mental accounting,” which points out that investors mentally separate their portfolios into subportfolios for which they have quite different utility functions or risk aversion parameters. For example, investors may have one set of portfolios that they never evaluate and another set that they evaluate every day. Benartzi and Thaler show that the size of the historical equity premium can be explained if investors evaluate their portfolio at least annually.


Credit default swap (CDS) spreads can be used to estimate the ERP because they are the directly observable market price of corporate performance risk, while equity expected returns are not directly observable and must be inferred. Their estimates of the ERP are high, ranging from 5% to over 7%, but they caution that, because of their assumptions, these should be regarded as upper limits, not best estimates.


By studying historical intervals when stock valuation (price/dividends or price/earnings) was the same at the end of the interval as at the beginning, one can avoid incorporating unexpected valuation changes into long-term rate of return studies. The analysis gives an equity risk premium of 3 percent, although the more interesting finding is that equity returns are mean-reverting whereas bond returns have no mean to which to regress. Thus, in the very long run and in real terms, stocks are safer than bonds.


This essential paper is best described in the authors’ abstract: “Two important concepts played a key role in the bull market of the 1990s. Both... are demonstrably untrue. First, many investors believed that earnings could grow faster than the macroeconomy. In fact, earnings must grow slower than GDP because the growth of existing enterprises contributes only part of GDP growth; ...the creation of new enterprises is [also] a key driver of GDP growth, and it does not contribute to the growth in earnings and dividends of existing enterprises. During the 20th century, growth in stock prices and dividends [not per share but in absolute magnitude] was 2% [per year] less than underlying macroeconomic growth.

Second, [for evaluating the impact of share buybacks on earnings per share], ...[t]he important metric is...net buybacks — stock buybacks less new share issuance...
We demonstrate...that, during the 20th century, new share issuance in many nations almost always exceeded stock buybacks by an average of 2 percent or more a year.


The authors trace the recent boom in share buybacks, as opposed to cash dividends, to a change in SEC rules in 1982 and find that dividends are managed conservatively (because there is a financial market penalty for cutting them) while buybacks are managed discretionarily (because there does not seem to be any such penalty).


The authors calculate returns on a global multi-asset portfolio, using fixed weights, and recommend it is a benchmark for pension plans. It is one of the first attempts to measure the return on world wealth, although the use of fixed weights means that the portfolio manager would have to transact in order to rebalance the these weights, instead of pursuing a buy-and-hold strategy. The use of fixed weights avoids the need to monitor the portfolio for excessive risk caused by one asset class becoming dominant over time.


This paper suggests that survival could induce a substantial spurious equity premium and at least partially explain the equity premium puzzle documented by Mehra and Prescott [1985]. (That is, to explain it away, because the returns used to frame the “puzzle” were neither expected nor were they achieved by many investors.)


In an extended literature review and personal essay, Campbell finds that “the world geometric average equity premium was almost 4% at the end of March 2007, implying a world arithmetic average equity premium somewhat above 5%. Both valuation ratios and the cross-section of stock prices imply that the equity premium fell considerably in the late 20th Century, but has risen modestly in the early years of the 21st Century.”


In what is almost certainly the first “CAPE” paper, the authors write that “a long moving average of real earnings helps to forecast future real dividends. The ratio of this earnings variable to the current stock price is a powerful predictor of the return on stock, particularly...over several years.” Thus, contrary to the efficient market hypothesis that prevailed in academia at the time this paper was written, stock prices are at least somewhat predictable and countercyclical, that is, expected returns are high when prices (relative to earnings and dividends) are low and low when prices are high.


The dividend-to-price ratio (D/P) can forecast either changes in dividend, which is what efficient market theory suggests, or changes in price, or both. Empirically, it forecasts only changes in price. At the then-current (1998) D/P, the forecast was extraordinarily bearish: The stock market was expected to lose about two-thirds of its real value. The forecast becomes less drastically bearish (although still quite bearish) when one uses (dividend + share buybacks), earnings, the 10-year moving average of earnings in
constant dollars, or other variables in the denominator. Real stock returns close to zero over the next 10 years were forecast. A number of statistical weaknesses in the analysis are acknowledged: The historical observations are not independent, and the analysis depends on valuation ratios regressing to their historical means, whereas the actual means are not known and could conceivably lie outside the historical range.


Presidential Address to the American Finance Association, focusing on changing beliefs in the field of finance. He indicates that “discount-rate variation, [including variation in the ERP] is the central organizing question of current asset-pricing research.” He describes the impact of research on discount-rate variation for “portfolio theory, accounting, cost of capital, capital structure, compensation, and macroeconomics.”


As the correlation of equities with personal income changes over the life of the investor, so does the attractiveness of equities to that investor. The young, who should borrow to smooth consumption and to invest in equities, cannot do so. Therefore, equities are priced almost exclusively by middle-aged investors, who find equities to be unattractive. The result is a decreased demand for equities and an increased demand for bonds relative to what it would be in a perfectly competitive market.


In an annually updated comprehensive review, Damodaran catalogues the various methods available for estimating the ERP, including “economic determinants of equity risk premiums…, the standard [historical] approach…, the survey approach…, and the implied approach, where a forward-looking estimate of the premium is estimated using either current equity prices or risk premiums in non-equity markets.” Damodaran also “look[s] at the relationship between the equity risk premium and risk premiums in the bond market…and in real estate…and [asks] how that relationship can be mined to generated expected equity risk premiums.”


In a sophisticated literature review and general discussion, the authors’ state reasons why others have found a “puzzle” or macroeconomic paradox in the high expected returns of equity markets and present accumulated evidence that equity returns have been and will continue to be high relative to riskless assets such as Treasury bills.


Stock total returns must equal dividend yields plus the growth rate of dividends, which cannot, in the long run, exceed the growth rate of the economy. If infinite-run expected dividend growth exceeded infinite-run expected economic growth, then dividends would crowd out all other economic claims. Net new issues, representing new capital (transferred from the labor market) that is needed so the corporate sector can grow, may cause the dividend growth rate to be slower than the GDP growth rate. Thus, the equity risk premium equals the dividend yield (minus new issues net of share buybacks), plus the GDP growth rate, minus the riskless rate.
The “supply side” or “supply model” thread begins with this work, which was written when so-called supply-side economics was popular; the authors strove to apply supply-oriented thinking to investment questions.


The authors document the annual returns on equities, bonds, bills, currencies, and inflation over the twentieth century for all major markets (the U.S., the U.K., Japan, France, Germany, Canada, Italy, Spain, Switzerland, Australia, the Netherlands, Sweden, Belgium, Ireland, Denmark, and South Africa). Later editions include more countries. The authors show that survival bias is a significant factor in estimating future returns, because past returns reflect only those countries that have been successful. In a speech, Dimson summarized his and his co-authors’ work as follows: “Although equities gave the highest return in every country, they were also risky, and we demonstrate the importance of diversifying globally as well as across asset classes.


Using 1900-2005 data for 17 countries, the authors show that the annualized equity premium for the rest of the world was 4.2 percent, not too much below the U.S. equity premium of 5.5 percent over the same period.

The historical equity premium is decomposed into dividend growth, multiple expansion, the dividend yield, and changes in the real exchange rate. Assuming zero change in the real exchange rate and no multiple expansion, and a dividend yield 0.5-1 percent lower than the historical mean (4.49 percent), the authors forecast a geometric equity premium on the world index around 3-3.5 percent and 4.5-5 percent on an arithmetic mean basis.


While the authors’ document geometric-mean realized equity risk premia ranging from 2.7% in Denmark to 6.7% in Australia over 1900-2009, they project lower returns going forward. Their estimate of the global ERP at the time of their writing was “3.4% for the average country and…4.0 percent for the World index.”


Annually updates Dimson, Marsh, and Staunton [2002].


These two Federal Reserve economists “categoriz[e] the [available] models into five classes: trailing historical mean, dividend discount, cross-sectional estimation, regression analysis using valuation ratios or macroeconomic variables, and surveys.” They make a one-year ERP forecast (as of June 2012) of 12.2%.


Annually updates Ibbotson and Sinquefield [1976].
Dividend yields predict intermediate- and long-horizon equity market returns much better than short-horizon returns. While regressions of returns on dividend yields typically explain less than 5% of monthly or quarterly return variances, the percentage explained rises to 25% to 40% for a three- to five-year horizon. This result sharply contradicts efficient markets and suggests that investors should buy when dividend yields are high and sell when they are low.


This paper compares alternative estimates of the unconditional expected stock return between 1872 and 2000, and explains the low expected return estimates derived from fundamentals, such as dividends and earnings, for the 1951-2000 period. The authors conclude that the decline in discount rates largely causes the unexplained capital gain of the last half-century.


This comprehensive volume includes the authors’ works, with many co-authors, from the 1970s to the 2000s. Because the authors produced much of the literature discussed in this Literature Review, Goetzmann and Ibbotson’s book of collected works is indispensable for serious scholars of the equity risk premium and related issues.


Survey-based measures of the returns investors expect are procyclical (they rise after markets have risen), while model-based estimates of expected returns are countercyclical (they fall after markets have risen). Thus, the returns investors say they expect are negatively correlated with the returns they would expect if they followed the (mostly DDM-based) models.


This is the predecessor article to Grinold, Kroner, and Siegel [2011] and provides some detail on the input estimation methods not included in the later article.


The authors examine the four components of the expected equity risk premium separately (income return, expected real earnings growth, expected inflation, and expected repricing) and suggest a then-current risk premium of about 3.6% over 10-year Treasury bonds. The authors also forcefully attack the “puzzle” literature, saying that they have never understood how one can seriously assert that the theory is right and the data are wrong.

A reconvening by the CFA Institute Research Foundation of the equity risk premium discussion group in Leibowitz et al. [2002], with some additions and deletions. Rather than a transcript of proceedings, this is a collection of articles submitted by the presenters after the meeting. Contributors, in alphabetical order by lead author, are Andrew Ang and Xiaoyan Zhang; Robert Arnott; Clifford Asness; Peng Chen; Elroy Dimson, Paul Marsh, and Mike Staunton; Richard Grinold, Kenneth Kroner, and Laurence Siegel; Brett Hammond and Martin Leibowitz; Roger G. Ibbotson; Antti Ilmanen; Rajnish Mehra; and Jeremy Siegel. Some of the individual articles are referenced separately in this essay.


The authors present annual return and market-capitalization data on global equities, global fixed income, commodity metals, and U.S real estate over 1960-1984. Cap-weighting the individual asset-class returns, they present a composite return series for the world market wealth portfolio. They note that what is omitted (human capital, non-U.S. real estate, private businesses) is probably larger than what is included.


Total equity returns consist of a stationary part (the equity risk premium) and a nonstationary part (the interest rate component, which consists of a real interest rate plus compensation for expected inflation). The estimator of the future arithmetic mean equity risk premium is the past arithmetic mean premium, which was about 7 percent when the authors wrote the article. To this is added the then-current interest rate, 4.8 percent (on 20-year Treasury bonds). The sum of these, about 12 percent, was the arithmetic mean expected total return on equities. The historical equity risk premium reflects equilibrium at all times and forms the proper estimator of the future equity risk premium. (Later updates discuss other methods rather than supporting a doctrinaire “future equals past” interpretation of historical data.)


The author takes a “cubic” approach to understanding expected returns. On one face of the cube are conventional asset classes: stocks, government bonds, credits, and alternatives. A second face considers trading strategies: value, “carry” (roughly speaking, yield), trend, and volatility. The third face is for underlying macroeconomic factors: growth, inflation, illiquidity, and tail risks. The treatment is encyclopedic and covers many aspects of return estimation, alpha generation, and beta-focused investing.


The author focuses on the difficulty of timing the market using time-varying valuation or risk premium approaches, but also engages in a very high-quality discussion of ERP issues in general.

The authors compare real stock returns over 1921-1996 in the U.S. with real stock returns in 38 other countries over the same period and find that the U.S. return was much higher. Thus survival bias is a significant factor in evaluating historical returns and the historical ERP. Simply projecting past returns forward into the future results in forecasts that are much too high.

http://www.cfapubs.org/toc/cp.1/2002/7

Transcript of a high-level discussion group convened by AIMR, now renamed the CFA Institute. Contributors include, in alphabetical order, Robert Arnott, Clifford Asness, Bradford Cornell, Campbell Harvey, Martin Leibowitz, Rajnish Mehra, Robert Shiller, Jeremy Siegel, and Richard Thaler.


The large run-up in equity value, relative to GDP between 1962 and 2000, is mainly caused by (1) large reductions in individual tax rates, (2) increased opportunities to hold equity in a nontaxed pension plan, and (3) increases in intangible and foreign capital. The authors argue that the high equity risk premium documented by Mehra and Prescott [1985] is not puzzling after these three factors are accounted for. However, in the future, one should expect no further gains from tax policy; the currently expected real return on equities is about 4 percent, down from 8 percent in the early postwar period.


The ERP puzzle literature is easily misunderstood because of its difficulty. Here, the puzzle is stated in language that is accessible to most finance practitioners. First, empirical facts regarding the returns and risks of major asset classes are presented. Then, the theory responsible for the “puzzle” is summarized. Modern asset pricing theory assumes that economic agents pursue and, on average, get fair deals. When one follows this line of reasoning to its conclusion, using the tools of classic growth and real business cycle theory, an equity risk premium of at most 1 percent emerges. An extensive discussion reveals why this is the case and addresses various attempts made by other authors to resolve the puzzle.


In this respected work, Mehra and Prescott first document the “equity premium puzzle” using a consumption-based asset-pricing model in which the quantity of risk is defined as the covariance of excess stock return with consumption growth and the price of risk is the coefficient of relative risk aversion. Because of the low risk resulting from the smooth historical growth of consumption, the 6 percent ERP in the 1889-1978 period can only be explained by a very high coefficient of risk aversion in the magnitude of 30 to 40. Risk aversion parameters observed in other aspects of financial behavior are around 1. Such a risk aversion parameter is consistent with at most a 1 percent equity risk premium, and possibly one as small as 0.25 percent.

Note that Mehra and Prescott assumed that consumption was equal to aggregate dividends. Because consumption is very smooth and dividends are not as smooth, this comparison may be troublesome.

Rietz suggests that the ERP puzzle can be solved by incorporating a very small probability of a very large drop in consumption. In such a scenario, the risk-free rate is much lower than the equity return. In an article published in the same issue of the Journal of Monetary Economics, Mehra and Prescott argued that Rietz’s model requires a 1 in 100 chance of a 25 percent decline in consumption to reconcile the equity premium with a risk aversion parameter of 10. However, the author says, the largest consumption decline in the last 100 years was only 8.8 percent.

But during the Great Depression, the stock market fell by 86 percent from peak to trough and dividends fell by about half; aggregate consumption in the economy, not just by stockholders, fell by about 18 percent. Mehra and Prescott’s 8.8 percent is the largest one-year decline in a multi-year consumption decline.


Irrational Exuberance, the title taken from an Alan Greenspan speech, presents basic concepts of behavioral finance and argues that markets become overextended so that returns can be above-normal, then below-normal, for extended periods.

Shiller introduces (to a mass audience) the concept of the CAPE, cyclically adjusted price-earnings ratio, which modifies the traditional PE ratio by using the average of 10 years’ trailing real earnings in place of (trailing or forecast) current year’s earnings. This method achieves a compromise, using a period longer than one year (to stabilize the earnings measure) but not too long (so that old, irrelevant data are not used). The ability of the CAPE to make market return forecasts is documented.


This immensely influential volume documents “Siegel’s constant” — the idea that real (inflation-adjusted) returns on stocks have been close to a constant over very long time periods. “Note the extraordinary stability of the real return on stocks,” the author writes, “over all major subperiods: 7.0 percent per year from 1802-1870, 6.6 percent from 1871 through 1925, and 7.2 percent per year since 1926.” Even in countries where stock markets were almost ruined by war, such as Germany and Japan, stocks beat bonds, which were entirely ruined in those countries.


While generally supportive of Shiller’s [2000] CAPE approach to market valuation, Siegel notes that accounting standards have become more conservative, notably around 1990 when depreciation requirements for goodwill changed. Thus CAPE ratios from before that time are not necessarily relevant to assessing the current valuation of the market. Using more contemporary data, the market appears less overvalued.

The author also recommends using National Income and Product Accounts (NIPA) earnings as a check on S&P 500 or other corporate-market earnings, because the former includes private businesses that are overlooked when the latter is used.


This is the predecessor article to the current one. Song covers most of the same issues as I do, but from a vantage point of a decade earlier, and with greater emphasis on the “puzzle” literature associated with Mehra and Prescott [1985].
The authors present evidence, over a 143-year period in the U.S., that total payouts (dividends plus buybacks), not dividends alone or earnings, are “the key drivers of long-run stock market returns.” They show that aggregate (not per share) total payouts grew at the same rate as GDP on average over time. The authors also introduce the CATY (cyclically adjusted total yield), that is, yield based on 10 years’ average real total payout, and show its predictive ability for returns.

This paper resolves a number of issues raised by Diermeier, Ibbotson, and Siegel [1984] and reconciles the DDM/DCF literature with Miller and Modigliani [1961], who show that investors should be indifferent between cash dividends and other forms of cash payout to the investor.
I'm not talking short term, long term or medium term - I just want to know if the lift is going 'up' or 'down'.