Volatility, Valuation Ratios, and Bubbles: An Empirical Measure of Market Sentiment

Can Gao       Ian Martin*

November, 2018

Abstract

We define a market sentiment indicator based on volatility and valuation ratios that exploits two contrasting views of return predictability, and study its properties. The indicator was unusually high during the late 1990s, reflecting dividend growth expectations that in our view were unreasonably optimistic; we interpret it as helping to reveal irrational beliefs about fundamentals. We also make two methodological contributions. First, we derive a new valuation-ratio decomposition that is related to the Campbell and Shiller (1988) loglinearization, but which resembles the traditional Gordon growth model more closely and has certain other advantages. Second, we introduce a market volatility index that provides a lower bound on the market’s expected log return.

*Can Gao: Imperial College, London, http://www.imperial.ac.uk/people/can.gao14 Ian Martin: London School of Economics, http://personal.lse.ac.uk/martiniw/ We thank John Campbell, Stefan Nagel, and seminar participants at Duke University and at the Federal Reserve Bank of New York for their comments. Ian Martin is grateful to the ERC for support under Starting Grant 639744.
This paper introduces a market sentiment indicator that exploits two contrasting views of market predictability.

A vast literature has studied the extent to which signals based on valuation ratios are able to forecast market returns and/or measures of dividend growth; early papers include Keim and Stambaugh (1986), Campbell and Shiller (1988), and Fama and French (1988). More recently, Martin (2017) argued that indexes of implied volatility based on option prices can serve as forecasts of expected excess returns; and noted that the two classes of predictor variables made opposing forecasts in the late 1990s, with valuation ratios pointing to low long-run returns and option prices pointing to high short-run returns.

Our paper trades off the two views of the world against one other. Consider the classic Gordon growth model, which relates the market’s dividend yield to its expected return minus expected dividend growth:

\[
\frac{D}{P} = \mathbb{E}(R - G).
\]

Very loosely speaking, the idea behind the paper is to use option prices to measure \(\mathbb{E}R\), and then to calculate the expected growth in fundamentals implicit in market valuations—our sentiment measure—as the difference between the option price index and dividend yield, \(\mathbb{E}G = \mathbb{E}R - \mathbb{E}(R - G)\).

Putting this thought into practice is not as easy as it might seem, however. For example, the Gordon growth model relies on assumptions that expected returns and expected dividend growth are constant over time. The loglinearized identity of Campbell and Shiller (1988) showed how to generalize the Gordon growth model to the empirically relevant case in which these quantities are time-varying. Their identity relates the price-dividend ratio of an asset to its expected future log dividend growth and expected log returns. It is often characterized as saying that high valuation ratios signal high expected dividend growth or low expected returns (or both). But expected returns are not the same as expected log returns. We show that high valuations—and low expected log returns—may be consistent with high expected returns if log returns are highly volatile, right-skewed, or fat-tailed. Plausibly, all of these conditions were satisfied in the late 1990s. As they are all potential explanations for the rise in valuation ratios at that time, we will need to be careful about the distinction between log returns and simple returns.

Furthermore, we show that while the Campbell–Shiller identity is highly accurate on average, the linearization is most problematic at times when the price-dividend ratio is far above its long-run mean. At such times—the late 1990s being a leading example—a researcher who uses the Campbell–Shiller loglinearization will conclude that long-run expected returns are even lower, and/or long-run expected dividend growth is even
higher, than is actually the case. We therefore propose a new loglinearization that does not have this feature, but which also relates a measure of dividend yield to expected log returns and dividend growth.

The second ingredient of our paper is a lower bound on expected log returns. (This plays the role of $\mathbb{E} R$ in the loose description above.) The lower bound relies on an assumption closely related to the negative correlation condition of Martin (2017); it can be computed directly from index option prices so is, broadly speaking, a measure of implied volatility.

Volatility and valuation ratios have, of course, long been linked to bubbles. A novel feature of our approach is that we use some theory to motivate our definitions of volatility and of valuation ratios, and to make the link quantitative. Our approach also satisfies the requirement noted by Brunnermeier and Oehmke (2013) that practically useful risk measures should be “measurable in a timely fashion.” There are various choices to be made regarding the details of the construction of the indicator: we have tried to make these choices in a conservative way to avoid “crying bubble” prematurely, in the hope that the indicator might be useful to cautious policymakers in practice.

The paper is organized as follows. Section 1 discusses the link between valuation ratios, returns, and dividend growth; it analyzes the properties of the Campbell–Shiller loglinearization, introduces our alternative loglinearization, and studies the predictive relationship between the dividend yield measures and future (log) returns and (log) dividend growth. Section 2 derives the lower bound on expected returns. Section 3 combines the preceding sections to introduce the sentiment indicator; then explores its relationship with volume and with (a measure of) the probability of a crash. Section 4 concludes.

1 Fundamentals

We seek to exploit the information in valuation ratios, following Campbell and Shiller (1988). We write $P_{t+1}$, $D_{t+1}$ and $R_{t+1}$ for the level, dividend, and gross return of the market, respectively: thus

$$R_{t+1} = \frac{D_{t+1} + P_{t+1}}{P_t}.$$  \hfill (1)

It follows from (1) that

$$r_{t+1} - g_{t+1} = pd_{t+1} - pd_t + \log \left(1 + e^{d_{t+1}}\right),$$  \hfill (2)
where we write \( dp_{t+1} = d_{t+1} - p_{t+1} = \log D_{t+1} - \log P_{t+1} \), \( pd_{t+1} = p_{t+1} - d_{t+1} \), and \( g_{t+1} = d_{t+1} - d_{t} \). Campbell and Shiller (1988) linearized the final term in (2) to derive a decomposition of the (log) price-dividend ratio

\[
pd_t = \frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t \left( g_{t+1+i} - r_{t+1+i} \right),
\]

where the constants \( k \) and \( \rho \) are determined by

\[
\rho = \frac{\mu}{1 + \mu} \quad \text{and} \quad \frac{k}{1 - \rho} = (1 + \mu) \log(1 + \mu) - \mu \log \mu, \quad \text{where} \quad \mu = e^{pd}.
\]

The approximation (3) is often loosely summarized by saying that high valuation ratios signal high expected dividend growth or low expected returns (or both). But expected log returns are not the same as expected returns\(^2\); we have

\[
\mathbb{E}_t r_{t+1+i} = \log \mathbb{E}_t R_{t+1+i} - \frac{1}{2} \text{var}_t r_{t+1+i} - \sum_{n=3}^{\infty} \frac{\kappa^{(n)}(r_{t+1+i})}{n!},
\]

where \( \kappa^{(n)}(r_{t+1+i}) \) is the \( n \)-th conditional cumulant of the log return. (If returns are conditionally lognormal, then the higher cumulants \( \kappa^{(n)}(r_{t+1+i}) \) are zero for \( n \geq 3 \).) Thus high valuations—and low expected log returns—may be consistent with high expected arithmetic returns if log returns are highly volatile, right-skewed, or fat-tailed. Plausibly, all of these conditions were satisfied in the late 1990s. As they are all potential explanations for the rise in valuation ratios at that time\(^3\) we will need to be careful about the distinction between log returns and simple returns.

Furthermore, the Campbell–Shiller first-order approximation is least accurate when the valuation ratio is far from its mean, as we now show.

**Result 1** (Campbell–Shiller revisited). The log price-dividend ratio \( pd_t \) obeys the fol-

\(^1\)We follow the convention in the literature in writing approximations such as (3) with equals signs. A number of our results below are in fact exact. We emphasize these as they occur. We also assume throughout the paper that there are no rational bubbles, as is standard in the literature. Thus, for example, in deriving (3) we are assuming that \( \lim_{T \to \infty} \rho_T^T pd_T = 0 \).

\(^2\)And expected log dividend growth is not the same as expected dividend growth. This distinction is less important, however, as log dividend growth is less volatile than log returns.

\(^3\)See, for example, Pastor and Veronesi (2003, 2006).
Following exact decomposition:

\[
pd_t = \frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i (g_{t+1+i} - r_{t+1+i}) + \frac{1}{2} \sum_{i=0}^{\infty} \rho^i \psi_{t+1+i} (1 - \psi_{t+1+i}) \left( pd_{t+1+i} - \overline{pd} \right)^2,
\]

where the constants \( k \) and \( \rho \) are defined as above, and the quantities \( \psi_{t+1+i} \) lie between \( \rho \) and \( 1/(1 + e^{dp_{t+1+i}}) \).

Equation (4) becomes a second-order Taylor approximation if \( \psi_t \) is assumed equal to \( \rho \) for all \( t \),

\[
pd_t = \frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i (g_{t+1+i} - r_{t+1+i}) + \frac{\rho(1 - \rho)}{2} \sum_{i=0}^{\infty} \rho^i \left( pd_{t+1+i} - \overline{pd} \right)^2,
\]

and reduces to the Campbell–Shiller loglinearization (3) if the final term on the right-hand side of (4) is neglected entirely.

Proof. Taylor’s theorem, with the Lagrange form of the remainder, states that (for any sufficiently well-behaved function \( f \), and for \( x \in \mathbb{R} \) and \( a \in \mathbb{R} \))

\[
f(x) = f(a) + (x - a) f'(a) + \frac{1}{2} (x - a)^2 f''(\xi),
\]

for some \( \xi \) between \( a \) and \( x \). (6)

We apply this result with \( f(x) = \log (1 + e^{x}) \), \( x = dp_{t+1} \), \( a = \overline{dp} = \mathbb{E} dp_t \) equal to the mean log dividend yield. Equation (6) becomes

\[
\log (1 + e^{dp_{t+1}}) = k + (1 - \rho) dp_{t+1} + \frac{1}{2} \psi_{t+1} (1 - \psi_{t+1}) \left( dp_{t+1} - \overline{dp} \right)^2,
\]

where \( \psi_{t+1} = 1/(1 + e^{\xi}) \) must lie between \( 1/(1 + e^{\overline{dp}}) = \rho \) and \( 1/(1 + e^{dp_{t+1}}) \).

Substituting into expression (2), we have the exact relationship

\[
r_{t+1} - g_{t+1} = k - pd_t + \rho pd_{t+1} + \frac{1}{2} \psi_{t+1} (1 - \psi_{t+1}) \left( pd_{t+1} - \overline{pd} \right)^2
\]

which can be solved forward to give the result (4). The approximation (5) follows.

Result (4) expresses the price-dividend ratio in terms of future log dividend growth and future log returns—as in the Campbell–Shiller approximation—plus a convexity correction.

This convexity correction is small on average. Take the unconditional expectation
of second-order approximation (5):

\[ \mathbb{E} \, pd_t = \frac{k}{1 - \rho} + \frac{\mathbb{E} (g_t - r_t)}{1 - \rho} + \frac{\rho}{2} \text{var} \, pd_t, \]

assuming that \( pd_t, r_t, \) and \( g_t \) are stationary so that their unconditional means and variances are well defined. Using CRSP data from 1947 to 2017, the sample average of \( pd_t \) is 3.483 (so that \( \rho \) is 0.970) and the sample standard deviation is 0.436. Thus the unconditional average convexity correction \( \frac{\rho}{2} \text{var} \, pd_t \) is about 0.0924, that is, about 2.65% of the size of \( \mathbb{E} \, pd_t \).

The convexity correction can be large conditionally, however. We have

\[ pd_t = \frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t (g_{t+1+i} - r_{t+1+i}) + \frac{\rho(1 - \rho)}{2} \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t (pd_{t+1+i} - \overline{pd})^2, \]

and the final term may be quantitatively important if the valuation ratio is far from its mean and persistent, so that it is expected to remain far from its mean for a significant length of time.

For the sake of argument, suppose the log price-dividend ratio follows an AR(1),

\[ pd_{t+1} - \overline{pd} = \phi(pd_t - \overline{pd}) + \varepsilon_{t+1}, \]

where \( \text{var} \, \varepsilon_{t+1} = \sigma^2 \) so that \( \text{var} \, pd_t = \sigma^2/(1-\phi^2) \); and set \( \sigma = 0.168 \) and \( \phi = 0.923 \) to match the sample standard deviation and autocorrelation in CRSP data from 1947–2017. The above expression becomes

\[ pd_t = \frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t (g_{t+1+i} - r_{t+1+i}) + \frac{\rho(1 - \rho)}{2(1 - \rho\phi^2)} \left[ (pd_t - \overline{pd})^2 + \frac{\sigma^2}{(1 - \rho)\phi^2} \right]. \]

At its peak during the boom of the late 1990s, \( pd_t \) was 2.2 standard deviations above its mean. The convexity term then equals 0.145: this is the amount by which a researcher using the Campbell–Shiller approximation would overstate \( \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t (g_{t+1+i} - r_{t+1+i}) \). With \( \rho = 0.970 \), this is equivalent to overstating \( \mathbb{E}_t g_{t+1+i} - r_{t+1+i} \) by 14.5 percentage points for one year, 3.1 percentage points for five years, or 1.0 percentage points for 20 years.

\[ ^4 \text{The numbers are more dramatic if we use the long sample from 1871–2015 available on Robert Shiller’s website. We find } \rho = 0.960, \sigma = 0.136, \text{ and } \phi = 0.942 \text{ in the long sample, so that the convexity correction is 0.0596 when } pd_t \text{ is at its mean, and 0.253 at the peak (which was 3.2 standard deviations above the mean). This last number correspond to overstatement } \mathbb{E}_t g_{t+1+i} - r_{t+1+i} \text{ by 25.3 percentage points for one year, 5.5 percentage points for five years, 1.8 percentage points for 20 years, or 1.0 percentage points for ever.} \]
The Campbell–Shiller approximation does not apply if \( dp_t \) follows a random walk (i.e., \( \mathbb{E}_t dp_{t+1} = dp_t \)). But in that case we can linearize \(^2\) around the conditional mean \( \mathbb{E}_t dp_{t+1} \) to find \(^5\)

\[
\mathbb{E}_t (r_{t+1} - g_{t+1}) = \log (1 + e^{dp_t}) = \log \left( 1 + \frac{D_t}{P_t} \right).
\] (7)

Motivated by this fact, we define \( y_t = \log (1 + D_t/P_t) \). An appealing property of this definition—and one that \( dp_t \) does not possess—is that \( y_t = \log (1 + D_t/P_t) \approx D_t/P_t \). We can then rewrite the definition of the log return \(^2\) as the (exact) relationship

\[
r_{t+1} - g_{t+1} = y_{t+1} + \log (e^{y_t} - 1) - \log (e^{y_{t+1}} - 1).
\] (8)

In these terms, equation (7) states that

\[
y_t = \mathbb{E}_t (r_{t+1} - g_{t+1}),
\] (9)

which is valid, as a first-order approximation, if \( dp_t \) (or \( y_t \)) follows a random walk.

Alternatively, if \( y_t \) is stationary (as is typically assumed in the literature) we have the following result. We write unconditional means as \( \overline{y} = \mathbb{E} y_t, \overline{r} = \mathbb{E} r_t \) and \( \overline{g} = \mathbb{E} g_t \).

**Result 2** (A variant of the Gordon growth model). We have the loglinearization

\[
y_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i (r_{t+1+i} - g_{t+1+i}),
\] (10)

where \(^6\) \( \rho = e^{-\overline{y}} \). As there is no constant in (10), and as \((1 - \rho) \sum_{i=0}^{\infty} \rho^i = 1\), this is a variant of the Gordon growth model: \( y \) is a weighted average of future \( r - g \).

To second order, we have the approximation

\[
y_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i (r_{t+1+i} - g_{t+1+i}) - \frac{1}{2} \frac{\rho}{1 - \rho} \sum_{i=0}^{\infty} \rho^i \left[(y_{t+1+i} - \overline{y})^2 - (y_{t+i} - \overline{y})^2\right].
\] (11)

We also have the exact relationship

\[
\overline{y} = \overline{r} - \overline{g},
\] (12)

\(^6\)Campbell (2008, 2018) derives the same result via a different route, under further assumptions (that the driving shocks are homoskedastic and conditionally Normal) that we do not require.

\(^6\)This differs slightly from the definition of \( \rho \) in Result 1, though they are extremely close in practice.
which does not rely on any approximation.

Proof. Using Taylor’s theorem to second order in equation (8), we have the second-order approximation

\[ r_{t+1} - g_{t+1} = \frac{1}{1 - \rho} y_t - \frac{\rho}{1 - \rho} y_{t+1} + \frac{\rho}{2 (1 - \rho)^2} [(y_{t+1} - \bar{y})^2 - (y_t - \bar{y})^2] \]

which can be rewritten

\[ y_t = (1 - \rho)(r_{t+1} - g_{t+1}) + \rho y_{t+1} - \frac{\rho}{2 (1 - \rho)} [(y_{t+1} - \bar{y})^2 - (y_t - \bar{y})^2] , \]

and then solved forward, giving (10) and (11). Equation (12) follows by taking expectations of the identity (8) and noting that \( E \log (e^{y_t} - 1) = E \log (e^{y_{t+1}} - 1) \) by stationarity of \( y_t \).

Given our focus on bubbles, we are particularly interested in the accuracy of these loglinearizations at times when valuation ratios are unusually high or, equivalently, when \( dp_t \) and \( y_t \) are unusually low. This motivates the following definition and result.

**Definition 1.** We say that \( y_t \) is far from its mean (at time \( t \)) if

\[ E_t [(y_{t+1+i} - \bar{y})^2] \leq (y_t - \bar{y})^2 \text{ for all } i \geq 0. \]  

**Example.**—If \( y_t \) follows an AR(1), then it is far from its mean if and only if it is at least one standard deviation from its mean.

**Result 3** (Signing the approximation errors). We can sign the approximation error in the Campbell–Shiller loglinearization (3):

\[ dp_t < -\frac{k}{1 - \rho} + \sum_{i=0}^{\infty} \rho^i E_t (r_{t+1+i} - g_{t+1+i}) . \]  

The first-order approximation (10) is exact on average. That is,

\[ E y_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i E (r_{t+1+i} - g_{t+1+i}) \]  

holds exactly, without any approximation. But if \( y_t \) is far from its mean then (up to a
second-order approximation)

\[ y_t \geq (1 - \rho) \sum_{i=0}^{\infty} \rho^i E_t (r_{t+1+i} - g_{t+1+i}). \]  \hspace{1cm} (16)

**Proof.** The inequality (14) follows immediately from (4) and equation (15) follows directly from equation (12). To establish the inequality (16), rewrite

\[ \sum_{i=0}^{\infty} \rho^i [ (y_{t+1+i} - \bar{y})^2 - (y_{t+i} - \bar{y})^2 ] = - (y_t - \bar{y})^2 + (1 - \rho) \sum_{i=0}^{\infty} \rho^i (y_{t+1+i} - \bar{y})^2 \]

\[ = (1 - \rho) \sum_{i=0}^{\infty} \rho^i [ (y_{t+1+i} - \bar{y})^2 - (y_t - \bar{y})^2 ] . \]  \hspace{1cm} (17)

The inequality then follows from (11), (13), and (17). \[ \square \]

Dividend yields, whether measured by \( dp_t \) or by \( y_t \), were unusually low around the turn of the millennium, indicating some combination of low future returns and high future dividend growth. Result 3 shows that an econometrician who uses the Campbell–Shiller approximation (3) at such a time—that is, who treats the inequality (14) as an equality—will overstate how low future returns, or how high future dividend growth, must be: and therefore may be too quick to conclude that the market is “bubbly.” In contrast, an econometrician who uses the approximation (10) will understate how low future returns, or how high future dividend growth, must be. Thus \( y_t \) is a conservative diagnostic for bubbles.

To place more structure on the relationship between valuation ratios and \( r \) and \( g \), we will make an assumption about the evolution of \( dp_t \) and \( y_t \) over time. The Campbell–Shiller approximation over one period states that \( r_{t+1} - g_{t+1} = k + dp_t - \rho dp_{t+1} \). If \( dp_t \) follows an AR(1) with autocorrelation \( \phi \) then \( E_t dp_{t+1} = \phi (dp_t - \bar{dp}) \), so

\[ E_t (r_{t+1} - g_{t+1}) = c + (1 - \rho \phi) dp_t, \]  \hspace{1cm} (18)

where we have absorbed constant terms into \( c \).

Conversely, the first-order approximation underlying Result 2 states that

\[ r_{t+1} - g_{t+1} = \frac{1}{1 - \rho} y_t - \frac{\rho}{1 - \rho} y_{t+1}. \]  \hspace{1cm} (19)
If $y_t$ follows an AR(1) with autocorrelation $\phi_y$ then this reduces to

$$E_t(r_{t+1} - g_{t+1}) = c + \frac{1 - \rho \phi_y}{1 - \rho} y_t,$$

where again we absorb constants into the intercept $c$. In view of (12), this can also be written without an intercept as

$$E_t(r_{t+1} - g_{t+1}) - (\bar{r} - \bar{g}) = \frac{1 - \rho \phi_y}{1 - \rho} (y_t - \bar{y}),$$

so that the deviation of $y_t$ from its long-run mean is proportional to the deviation of conditionally expected $r_{t+1} - g_{t+1}$ from its long-run mean. A further advantage of $y_t$ over $dp_t$ is that the expression (20) is also meaningful if $y_t$ follows a random walk: in this case, the coefficient on $y_t$ equals one and the intercept is zero, by equation (9).

Equations (18) and (20) motivate regressions of realized $r_{t+1} - g_{t+1}$ onto $dp_t$ and a constant, or onto $y_t$ and a constant. The results are shown in Table 1; we also report the results of regressing $r_{t+1}$ and $-g_{t+1}$ separately onto $y_t$ and onto $dp_t$. We use end-of-year observations of the price level and accumulated dividends of the S&P 500 index.
from CRSP. The table reports regression results in the form

\[ \text{LHS}_{t+1} = a_0 + a_1 \times \text{RHS}_t + \varepsilon_{t+1}, \]

with Hansen–Hodrick standard errors. (Under the AR(1) assumption, we could also use (18) or (20) as estimates of \( \mathbb{E}_t(r_{t+1} - g_{t+1}) \). This approach turns out to give very similar results, as we show in the appendix: see Table 11.)

The variables \( y_t \) and \( dp_t \) have similar predictive performance and, consistent with the prior literature, we find, in the post-1947 sample, that valuation ratios help to forecast returns but have limited forecasting power for dividend growth. Table 2 reports results using cash reinvested dividends in the post-1926 period, which is the longest sample CRSP has. Tables 3 and 4 report similar results using semi-annual data. Tables 5 to 8 report results using the NYSE value-weighted index price and dividend data and compare them with market reinvested S&P500 data. Table 9 uses the price and dividend data of Goyal and Welch (2008) (updated to 2017 and taken from Amit Goyal’s webpage): this gives us a longer sample, as it incorporates Robert Shiller’s data which goes back as far as 1871. The predictability of \( r \) relative to \( g \) is to some extent a feature of the post-war period. In the long sample, returns are substantially less predictable and dividends substantially more predictable, perhaps because of the post-war tendency of corporations to smooth dividends (Lintner 1956). Encouragingly, though, we find that the predictive relationship between \( y_t \) (or \( dp_t \)) and the difference \( r_{t+1} - g_{t+1} \) is fairly stable across sample periods and data sources.

2 A lower bound on expected log returns

High valuation ratios are sometimes cited as direct evidence of a bubble. But valuation ratios can be high for good reasons if interest rates or rationally expected risk premia are low. In other words, if we use \( y_t \) to measure \( \mathbb{E}_t(r_{t+1} - g_{t+1}) \) as suggested above, we may find that \( y_t \) is low simply because \( \mathbb{E}_t r_{t+1} \) is very low, which could reflect low interest rates \( r_{f,t+1} \), low (log) risk premia \( \mathbb{E}_t r_{t+1} - r_{f,t+1} \), or both.

\(^7\)We calculate the monthly dividend by multiplying the difference between monthly cum-dividend and ex-dividend returns by the lagged ex-dividend price: \( D_t = (R_{\text{cum},t} - R_{\text{ex},t})P_{t-1} \). As we aggregate the dividends paid out over the year, to address seasonality issues, we reinvest dividends month-by-month until the end of the year, using the CRSP 30-day T-bill rate as our risk-free rate. In the appendix, we report similar results with dividends reinvested at the cum-dividend market return rather than at a risk-free rate; if anything, these results are somewhat more favorable to our \( y_t \) variable than to \( dp_t \).
While interest rates are directly observable, risk premia are harder to measure. We start from the following identity, which generalizes an identity introduced by Martin (2017) in the case $X_{t+1} = R_{t+1}$:

$$E_t X_{t+1} = \frac{1}{R_{f,t+1}} E_t^* (R_{t+1} X_{t+1}) - \text{cov}_t (M_{t+1} R_{t+1}, X_{t+1}).$$

We have written $E_t^*$ for the time-$t$ conditional risk-neutral expectation operator, defined by the property that $\frac{1}{R_{f,t+1}} E_t^* X_{t+1} = E_t (M_{t+1} X_{t+1})$ for any tradable payoff $X_{t+1}$ received at time $t+1$. Assuming the absence of arbitrage, the identity holds if the payoff $R_{t+1} X_{t+1}$ is tradable; it applies for any stochastic discount factor $M_{t+1}$ and gross return $R_{t+1}$, though for our purposes $R_{t+1}$ will always be the gross return on the market. We are interested in expected log returns, $X_{t+1} = \log R_{t+1}$, in which case the identity becomes

$$E_t \log R_{t+1} = \frac{1}{R_{f,t+1}} E_t^* (R_{t+1} \log R_{t+1}) - \text{cov}_t (M_{t+1} R_{t+1}, \log R_{t+1}). \tag{21}$$

To make further progress, we make two assumptions. As we will see below, we will use option prices to bound the first term on the right-hand side of the identity (21). Our first assumption addresses the minor technical issue that we observe options on the ex-dividend value of the index, $P_{t+1}$, rather than on $P_{t+1} + D_{t+1}$.

**Assumption 1.** If we define the dispersion measure

$$\Psi(X_{t+1}) \equiv E_t^* f(X_{t+1}) - f(E_t^* X_{t+1}),$$

where $f(x) = x \log x$ is a convex function, then the dispersion of $R_{t+1}$ is at least as large as that of $P_{t+1}/P_t$:

$$\Psi(R_{t+1}) \geq \Psi(P_{t+1}/P_t). \tag{22}$$

This condition is very mild. Expanding $f(x) = x \log x$ as a Taylor series to second order around $x = 1$, $f(x) \approx (x^2 - 1)/2$. Thus, to second order, Assumption 1 is equivalent to $\text{var}_t^* R_{t+1} \geq \text{var}_t^* (P_{t+1}/P_t)$, or equivalently $\text{var}_t^* (P_{t+1} + D_{t+1}) \geq \text{var}_t^* P_{t+1}$. A sufficient, though not necessary, condition for this to hold is that the price $P_{t+1}$ and dividend $D_{t+1}$ are weakly positively correlated under the risk-neutral measure.

Our second assumption is more substantive.

---

8In fact, it is so minor that the distinction between options on $P_{t+1}$ and options on $P_{t+1} + D_{t+1}$ is often neglected entirely in the literature. For example, Neuberger (2012) “avoid[s] irrelevant complications with interest rates and dividends” by treating options on forward prices as observable, as do Schneider and Trojani (2018), and (essentially equivalently) Carr and Wu (2009) use options on stocks as proxies for options on stock futures. The analogous assumption for our purposes would be that inequality (22) holds with equality.
Assumption 2. The modified negative correlation condition holds:

\[
\text{cov}_t (M_{t+1} R_{t+1}, \log R_{t+1}) \leq 0.
\] (23)

Martin (2017) imposed the closely related negative correlation condition (NCC) that \(\text{cov}_t (M_{t+1} R_{t+1}, R_{t+1}) \leq 0\). The two conditions are equivalent in the lognormal case, as we show below, and more generally the two are plausible for similar reasons: in any reasonable model, \(M_{t+1}\) will be negatively correlated with the return on the market, \(R_{t+1}\), and we know from the bound of Hansen and Jagannathan (1991), coupled with the empirical fact that high Sharpe ratios are available, that \(M_{t+1}\) is highly volatile.

The following two examples are adapted from Martin (2017).

Example 1.—Suppose that the SDF \(M_{t+1}\) and return \(R_{t+1}\) are conditionally jointly lognormal and write \(r_{f,t+1} = \log R_{f,t+1}\), \(\mu_t = \log \mathbb{E}_t R_{t+1}\), and \(\sigma_t^2 = \text{var}_t \log R_{t+1}\). Then the modified NCC is equivalent to the assumption that the conditional Sharpe ratio of the asset, \(\lambda_t \equiv (\mu_t - r_{f,t+1})/\sigma_t\), exceeds its conditional volatility, \(\sigma_t\); and hence also equivalent to the original NCC, \(\text{cov}_t (M_{t+1} R_{t+1}, R_{t+1}) \leq 0\).

Proof. By Stein’s lemma, \(\text{cov}_t (M_{t+1} R_{t+1}, \log R_{t+1}) = \text{cov}_t (\log M_{t+1} + \log R_{t+1}, \log R_{t+1})\). By lognormality of \(M_{t+1}\) and \(R_{t+1}\), the fact that \(\mathbb{E}_t (M_{t+1} R_{t+1}) = 1\) is equivalent to \(\log \mathbb{E}_t M_{t+1} + \log \mathbb{E}_t R_{t+1} = - \text{cov}_t (\log M_{t+1}, \log R_{t+1})\). It follows from these two facts that \(\text{cov}_t (M_{t+1} R_{t+1}, \log R_{t+1}) \leq 0\) if and only if \(\text{var}_t \log R_{t+1} \leq \log \mathbb{E}_t R_{t+1} - r_{f,t+1}\); that is, if and only if \(\lambda_t \geq \sigma_t\). This condition is equivalent to \(\text{cov}_t (M_{t+1} R_{t+1}, R_{t+1}) \leq 0\) in the lognormal case, as shown by Martin (2017).

Thus the modified NCC holds in the models of Campbell and Cochrane (1999), Bansal and Yaron (2004), Bansal et al. (2014) and Campbell et al. (2016), among many others.

Our second example does not require lognormality.

Example 2.—Suppose that there is an unconstrained investor who maximizes expected utility over next-period wealth, who chooses to invest his or her wealth fully in the stock market, and whose relative risk aversion (which need not be constant) is at least one at all levels of wealth. Then the modified NCC holds for the market return.

Proof. The given conditions imply that the SDF is proportional (with a constant of proportionality that is known at time \(t\)) to \(u'(W_t R_{t+1})\). We must therefore show that \(\text{cov}_t (u'(W_t R_{t+1}) R_{t+1}, \log R_{t+1}) \leq 0\). This holds for the very strong reason—much stronger than is actually needed for the NCC or modified NCC to hold—that
\[ u'(W_t R_{t+1}) R_{t+1} \] is decreasing in \( R_{t+1} \): its derivative is \[ u'(W_t R_{t+1}) + W_t R_{t+1} u''(W_t R_{t+1}) = -u'(W_t R_{t+1}) [\gamma(W_t R_{t+1}) - 1], \] which is negative because relative risk aversion \( \gamma(x) \equiv -x u''(x)/u'(x) \) is at least one.

We can now state our lower bound on expected log returns.

**Result 4.** Suppose Assumptions 1 and 2 hold. Write \( \text{call}_t(K) \) and \( \text{put}_t(K) \) for the time \( t \) prices of call and put options on \( P_{t+1} \) with strike \( K \), and \( F_t \) for the time \( t \) forward price of the index for settlement at time \( t + 1 \). Then we have

\[
\mathbb{E}_t r_{t+1} - r_{f,t+1} \geq \frac{1}{P_t} \left\{ \int_0^{F_t} \frac{\text{put}_t(K)}{K} dK + \int_{F_t}^{\infty} \frac{\text{call}_t(K)}{K} dK \right\}. \tag{24}
\]

**Proof.** As \( \mathbb{E}_t^* R_{t+1} = R_{f,t+1} \) and \( \mathbb{E}_t^* P_{t+1} = F_t \), the inequality (22) can be rearranged as

\[
\frac{1}{R_{f,t+1}} \mathbb{E}_t^* R_{t+1} \log R_{t+1} - \log R_{f,t+1} \geq \frac{1}{R_{f,t+1}} \left[ \mathbb{E}_t^* \left( \frac{P_{t+1}}{P_t} \log \frac{P_{t+1}}{P_t} \right) - \frac{F_t}{P_t} \log \frac{F_t}{P_t} \right]. \tag{25}
\]

The right-hand side of this inequality can be measured directly from option prices using a result of Breeden and Litzenberger (1978), which can be rewritten to give, for any sufficiently well behaved function \( g(\cdot) \),

\[
\frac{1}{R_{f,t+1}} \mathbb{E}_t^* g(P_{t+1}) - g(\mathbb{E}_t^* P_{t+1}) = \int_0^{F_t} g''(K) \text{put}_t(K) dK + \int_{F_t}^{\infty} g''(K) \text{call}_t(K) dK.
\]

Setting \( g(x) = \frac{x}{P_t} \log \frac{x}{P_t} \), we have \( g''(x) = 1/(P_t x) \). Thus

\[
\frac{1}{R_{f,t+1}} \left[ \mathbb{E}_t^* \left( \frac{P_{t+1}}{P_t} \log \frac{P_{t+1}}{P_t} \right) - \frac{F_t}{P_t} \log \frac{F_t}{P_t} \right] = \frac{1}{P_t} \left\{ \int_0^{F_t} \frac{\text{put}_t(K)}{K} dK + \int_{F_t}^{\infty} \frac{\text{call}_t(K)}{K} dK \right\}. \tag{26}
\]

The result follows on combining the identity (21), the inequalities (23) and (25), and equation (26).

We refer to the right-hand side of equation (24) as LVIX because it is reminiscent of the definition of the VIX index which, in our notation, is

\[
\text{VIX}_t^2 = 2 R_{f,t+1} \left\{ \int_0^{F_t} \frac{\text{put}_t(K)}{K^2} dK + \int_{F_t}^{\infty} \frac{\text{call}_t(K)}{K^2} dK \right\},
\]
and of the SVIX index introduced by Martin (2017),

$$\text{SVIX}_t^2 = \frac{2}{R_{f,t+1} F_t^2} \left\{ \int_0^{F_t} \text{put}_t(K) \, dK + \int_{F_t}^{\infty} \text{call}_t(K) \, dK \right\}. $$

We do not annualize our definition (24), so to avoid unnecessary clutter we have also not annualized the definitions of VIX and SVIX above. We will typically choose the period length from \( t \) to \( t + 1 \) to be six or 12 months. The forecasting horizon dictates the maturity of the options, so for example we use options expiring in six months to measure expectations of six-month log returns.

VIX, SVIX, and LVIX place differing weights on option prices. VIX has a weighting function \( 1/K^2 \) on the prices of options with strike \( K \); LVIX has weighting function \( 1/K \); and SVIX has a constant weighting function. In this sense we can think of LVIX as lying half way between VIX and SVIX. (We could also introduce a factor of two into the definition of LVIX to make the indices look even more similar to one another, but have chosen not to.)

We calculate LVIX using end-of-month interest rates and S&P 500 index option prices from OptionMetrics; full details of the calculation are provided in the appendix. Figure 1 plots \( \text{LVIX}_t \) over our sample period from January 1996 to December 2017.

2.1 A benchmark case

It is natural—both from an empirical perspective, but also as a guide to intuition—to wonder whether the inequality (24) might (approximately) hold with equality. For
this to be the case, we would need both (22) and (23) to hold with (approximate) equality. As the conditional volatility of dividends is substantially lower than that of prices, it is reasonable to think that this is indeed the case for (22), and as noted in footnote 8 much of the literature implicitly makes that assumption. Meanwhile the modified NCC (23) would hold with equality if one thinks from the perspective of an investor with log utility who chooses to hold the market, as is clear from the proof provided in Example 2 above. (The perspective of such an investor has been shown to provide a useful benchmark for forecasting returns on the stock market (Martin, 2017), on individual stocks (Martin and Wagner, 2018), and on currencies (Kremens and Martin, 2018).)

Table 10 in the Appendix reports the results of running the regression

\[ r_{t+1} - r_{f,t+1} = \alpha + \beta \times \text{LVIX}_t + \varepsilon_{t+1} \]  

(27)

at horizons of 3, 6, 9, and 12 months. Returns are computed by compounding CRSP monthly gross return of S&P 500. We report Hansen–Hodrick standard errors to allow for heteroskedasticity and autocorrelation that arises due to overlapping observations. If the inequality (24) holds with equality, we should find \( \alpha = 0 \) and \( \beta = 1 \). We do not reject this hypothesis at any horizon; and at the six- and nine-month horizons we can reject the hypothesis that \( \beta = 0 \) at conventional significance levels.

### 3 A sentiment indicator

We can now put the pieces together. We will measure expectations about fundamentals by subtracting \( \mathbb{E}_t(r_{t+1} - g_{t+1}) \), as revealed by valuation ratios under our AR(1) assumption, from \( \mathbb{E}_t r_{t+1} \), as revealed by interest rates and option prices:

\[
\mathbb{E}_t g_{t+1} = r_{f,t+1} + \mathbb{E}_t (r_{t+1} - r_{f,t+1}) - \mathbb{E}_t (r_{t+1} - g_{t+1}) \\
\geq r_{f,t+1} + \text{LVIX}_t - \mathbb{E}_t (r_{t+1} - g_{t+1}).
\]  

(28)

The inequality follows (under our maintained Assumptions 1 and 2) because \( \mathbb{E}_t r_{t+1} - r_{f,t+1} = \text{LVIX}_t \), as shown in Result 4.

We refer to the lower bound as the sentiment indicator, \( B_t \). Our central definition uses \( y_t \) to measure \( \mathbb{E}_t(r_{t+1} - g_{t+1}) \) via the fitted value \( \hat{a}_0 + \hat{a}_1 y_t \), as in Table 11 giving

\[
B_t = r_{f,t+1} + \frac{1}{P_t} \left[ \int_0^{F_t} \frac{\text{put}_t(K)}{K} dK + \int_{F_t}^{\infty} \frac{\text{call}_t(K)}{K} dK \right] - (\hat{a}_0 + \hat{a}_1 y_t).
\]
We estimate the coefficients $\hat{a}_0$ and $\hat{a}_1$ on a rolling basis: for example, at time $t$ they are estimated using data from 1947 until time $t$. Thus $B_t$ is observable at time $t$.

If $\mathbb{E}_t g_{t+1}$ itself follows an AR(1), as in the work of Bansal and Yaron (2004) and many others, then $B_t$ can also be interpreted as a (rescaled) lower bound on long-run dividend expectations. For if we have $\mathbb{E}_{t+1} g_{t+2} - \bar{g} = \phi_g (\mathbb{E}_t g_{t+1} - \bar{g}) + \varepsilon_{g,t+1}$ then long-run expected dividend growth at time $t$ is\footnote{We introduce the factor $1 - \rho$ in the definition of long-run expected dividend growth so that the weights $(1 - \rho)\rho^i$ sum to 1 and long-run expected dividend growth can be interpreted as a weighted average of all future periods’ expected growth.}

$$
(1 - \rho) \sum_{i \geq 0} \rho^i (\mathbb{E}_t g_{t+1+i} - \bar{g}) = \frac{1 - \rho}{1 - \rho \phi_g} (\mathbb{E}_t g_{t+1} - \bar{g}).
$$

The left panel of Figure 2 plots $B_t$ over our sample period\footnote{Figure 6 in the appendix, shows the corresponding results using the full sample period from 1947 to 2017 to estimate the relationship between $y_t$ (or $dp_t$) and $r_{t+1} - g_{t+1}$; but these time series were not observable in real time.}. The figure also plots a modified indicator, $B_{dp,t}$, that uses $dp_t$ rather than $y_t$ to measure $\mathbb{E}_t (r_{t+1} - g_{t+1})$ (as in (18)). This has the advantage of familiarity—$dp_t$ has been widely used in the literature—but the disadvantage that it may err on the side of signalling a bubble too soon, as shown in Result 3. Consistent with this prediction, the two series line up fairly closely, but $B_{dp,t}$ is less conservative—in that it suggests even higher $\mathbb{E}_t g_{t+1}$—during the period in the late 1990s when valuation ratios were far from their mean.

Note, moreover, that net dividend growth satisfies $\mathbb{E}_t \frac{D_{t+1}}{D_t} - 1 > \mathbb{E}_t g_{t+1}$, because $e^{g_{t+1}} > 1 + g_{t+1}$. Thus our lower bound on expected log dividend growth implies still
higher expected arithmetic dividend growth. As a rule of thumb, if dividend growth is conditionally lognormal then we would have log $E_t \frac{D_{t+1}}{D_t} = E_t g_{t+1} + \frac{1}{2} \text{var}_t g_{t+1}$. The variance term is small unconditionally—we have $\text{var} g_{t+1} \approx 0.005$ in our sample period—but it is plausible that during the late 1990s there was unusually high uncertainty about log dividend growth.

The right panel of Figure 2 plots the three components of the bubble indicator $B_t$ from 1996 to 2018. LVIX and $E_t(g_{t+1} - r_{t+1})$ moved in opposite directions for most of our sample period, with high valuation ratios occurring at times of low risk premia. But all three components were above their mean during the late 1990s.

3.1 The relationship with volume

It is often argued that high volume is a signature of bubbles (see, for example, Harrison and Kreps, 1978; Duffie, Gárateanu and Pedersen, 2002; Cochrane, 2003; Lamont and Thaler, 2003; Ofek and Richardson, 2003; Scheinkman and Xiong, 2003; Hong, Scheinkman and Xiong, 2006; Barberis et al., 2018). We construct a daily measure of volume using Compustat data from January 1983 to December 2017, by summing the product of shares traded and daily low price over all S&P 500 stocks on each day. As volume trended strongly upward during our sample period, we subtract a linear trend from log volume. We do so on a rolling basis using backward looking data, so our detrended volume measure at time $t$ is (like $B_t$) observable at time $t$. Figure 3a plots detrended log volume and $B_t$ over the sample period.

As is clear from the figure, our bubble measure is a leading indicator of detrended volume. Figure 3b plots the correlation between $B_{t+k}$ and detrended log volume at time $t$ for a range of values of $k$. Units are measured in months, so the peak at $k = -10$ months indicates that the correlation between the bubble index at time $t - 10$ months and detrended log volume at time $t$ is just over 91%. (Figure 7 in the appendix, shows plots that use full-sample, as opposed to real-time, information to compute both $B_t$ and the detrended volume measure.)

---

11We also constructed the corresponding measure using daily high prices: this gives essentially identical results.
3.2 The probability of a crash

One expects that the probability of a crash should be higher during a bubble episode. If not, the episode is perhaps not actually a bubble. We compare our bubble indicator with the crash probability measure derived by Martin (2017, Result 2). More precisely, we can calculate the probability of a market decline of a fixed percentage over a fixed period of time, as perceived by a log investor who holds the market, from option prices via the expression

$$P(R_{t+1} < \alpha) = \alpha \left[ \frac{\text{put}'(\alpha P_t)}{\alpha P_t} - \frac{\text{put}(\alpha P_t)}{\alpha P_t} \right]$$

(29)

where put'(K) is the first derivative of put price as a function of strike, evaluated at K.

The probability of a crash is high when out-of-the-money put prices are highly convex, as a function of strike, at strikes at and below $\alpha P_t$. By contrast, the measure of volatility that is relevant for our bubble indicator is sensitive to option prices across the full range of strikes of out-of-the-money puts and calls.

Figure 4 plots the two measures over time. The probability of a crash was elevated during the late 1990s, consistent with standard intuition about bubbles. But the crash probability was also high in the aftermath of the subprime crisis, an episode that we would certainly not identify as bubbly.

Figure 4: The bubble indicator and the probability of a 20% decline in the market over the next six months, as perceived by the log investor.

3.3 What if this time really is different?

A skeptic might argue that our measure of $\mathbb{E}_t(r_{t+1} - g_{t+1})$, which is based on an assumption that $y_t$ (or $dp_t$) follows an AR(1), breaks down during the late 1990s. If the breakdown is assumed to be temporary—a brief period during which valuation ratios behave differently, before subsequently reverting to business-as-usual—then we are happy to absorb such an interpretation into our definition of a bubble.

But what if one were prepared to believe in a genuinely New Economy? An aggressive skeptic might argue that the price-dividend ratio had ceased to mean-revert entirely. Conversely, a cautious central banker might justify inaction on the basis that valuation ratios could remain very high indefinitely.

Either perspective suggests considering the possibility that valuation ratios follow a random walk, $pd_t = \mathbb{E}_t pd_{t+1}$. If so, then $y_t = \mathbb{E}_t(r_{t+1} - g_{t+1})$ from equation (9), so that

$$\mathbb{E}_t g_{t+1} = \mathbb{E}_t r_{t+1} - y_t \geq \text{LVIX}_t + r_{f,t} - y_t.$$ 

This motivates the indicator $\tilde{B}_t = \text{LVIX}_t + r_{f,t} - y_t$. Aside from its appeal to the conservative central banker, this has the further advantage of not requiring estimation of any parameters. Figure 5 plots the time series of $\tilde{B}_t$ against the real-time version of $B_t$. Even if valuation ratios were expected to follow a random walk in the late 1990s—

---

\textsuperscript{13}As mentioned in footnote 5, Campbell (2008, 2018) considered this possibility—and perhaps with the same motivation.
a dubious proposition in any case—the implied expectations about cashflow growth appear implausibly high.

We note that $\tilde{B}_t$ also has a natural interpretation if $y_t$ follows an AR(1). For if dividend growth is unforecastable, as in the work of Campbell and Cochrane (1999) among many others, then valuation ratios directly reveal long-run expectations of log returns while LVIX reveals the corresponding short-run expectation, so the loglinearization (10) and the inequality $E_t r_{t+1} - r_{f,t} \geq \text{LVIX}_t$ together imply, after some algebra\footnote{See details in appendix D.1} that

$$E_t r_{t+1} - (1 - \rho) \sum_{i \geq 0} \rho^i E_t r_{t+2+i} \geq \frac{\tilde{B}_t - \bar{y}}{\rho}.$$  

Under this interpretation, $\tilde{B}_t$ is a rescaled measure of short-run expected log returns relative to subsequent long-run expected log returns.

## 4 Conclusion

We have presented a lower bound on expected dividend growth that exploits information in interest rates, index option prices, and the market valuation ratio. The lower bound was extraordinarily high during the late 1990s, reflecting dividend growth expectations that in our view were unreasonably optimistic. We therefore interpret it as
a sentiment indicator that helps to reveal irrational beliefs about fundamentals.

In simple terms, we characterize the late 1990s as a bubble because valuation ratios were high and short-run expected returns—as revealed by interest rates and our LVIX measure of implied volatility—were also high. Both aspects are important. We would not view high valuation ratios at a time of low expected returns, or low valuation ratios at a time of high expected returns, as indicative of a bubble (on the contrary, the latter scenario occurs in the aftermath of the market crash in 2008).

Our measure does not point to an unreasonable level of market sentiment in recent years, for example, as it interprets high valuation ratios as being justified by the low level of interest rates and implied volatility. A skeptic might respond that the low level of implied volatility is itself indicative of unreasonable complacency. This is, in principle, a possibility. Valuation ratios, interest rates and volatility could be internally consistent in our sense—so that our measure would not signal that anything is amiss—while also being mispriced. Our approach should be viewed as a test of the internal coherence of valuation ratios, interest rates, and option prices, rather than as a panacea.

It might seem strange that we rely on asset prices to provide a rational lower bound on expected log returns (via Result 4) while simultaneously arguing that the market itself was mispriced during part of our sample period. To sharpen the point, consider the special case discussed in Section 2.1. Our volatility measure LVIX then directly measures the expected excess log return perceived by a rational log investor who chooses to hold the market. Yet we simultaneously claim that there was a bubble in the late 1990s. These positions may appear to be inconsistent—why would a rational investor hold an overvalued stock market?—but they are not. As shown by the Campbell–Shiller loglinearized identity and by our variant, one can simultaneously have high short-run expected log returns $E_t r_{t+1}$, high valuation ratios $pd_t$, and make rational forecasts of fundamentals $\sum_{i=0}^{\infty} \rho^i E_t g_{t+1+i}$, so long as future expected returns $\sum_{i>0} \rho^i E_t r_{t+1+i}$ are low. And, critically, the log investor does not care about expected returns in future: he or she is myopic, so can be induced to hold the market by high short-run returns $E_t r_{t+1}$ whatever his or her beliefs about subsequent expected returns.

That said, for the investor’s expected log returns to be consistent with rationally expected log dividend growth during the bubble period, he or she must—given inequality—have believed that the historical forecasting relationship between dividend yield and $E_t (r_{t+1} - g_{t+1})$ had broken down. This is equivalent (by Result 18 or Result 20) to believing that dividend yields had ceased to mean-revert in the AR(1) manner suggested by prior history. This strikes us as a reasonable viewpoint for a rational investor living through
a bubble. It is consistent with the findings of Brunnermeier and Nagel (2004), who argued that in the late 1990s sophisticated investors such as hedge funds positioned themselves to exploit high short-run returns despite being skeptical about longer run returns, and with the view of the world colorfully articulated by former Citigroup chief executive Chuck Prince in a July, 2007, interview with the Financial Times: “When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you’ve got to get up and dance. We’re still dancing.”
References


# A Regression tables

<table>
<thead>
<tr>
<th>RHS$_t$</th>
<th>LHS$_{t+1}$</th>
<th>$\hat{a}_0$</th>
<th>s.e.</th>
<th>$\hat{a}_1$</th>
<th>s.e.</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t$</td>
<td>$r_{t+1} - g_{t+1}$</td>
<td>$-0.142$</td>
<td>$[0.054]$</td>
<td>$5.064$</td>
<td>$[1.507]$</td>
<td>$12.98%$</td>
</tr>
<tr>
<td></td>
<td>$r_{t+1}$</td>
<td>$0.032$</td>
<td>$[0.052]$</td>
<td>$1.694$</td>
<td>$[1.371]$</td>
<td>$2.04%$</td>
</tr>
<tr>
<td></td>
<td>$-g_{t+1}$</td>
<td>$-0.174$</td>
<td>$[0.051]$</td>
<td>$3.370$</td>
<td>$[1.502]$</td>
<td>$16.31%$</td>
</tr>
<tr>
<td>$dp_t$</td>
<td>$r_{t+1} - g_{t+1}$</td>
<td>$0.549$</td>
<td>$[0.187]$</td>
<td>$0.150$</td>
<td>$[0.053]$</td>
<td>$9.11%$</td>
</tr>
<tr>
<td></td>
<td>$r_{t+1}$</td>
<td>$0.321$</td>
<td>$[0.156]$</td>
<td>$0.067$</td>
<td>$[0.045]$</td>
<td>$2.59%$</td>
</tr>
<tr>
<td></td>
<td>$-g_{t+1}$</td>
<td>$0.229$</td>
<td>$[0.163]$</td>
<td>$0.082$</td>
<td>$[0.046]$</td>
<td>$7.84%$</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>RHS$_t$</th>
<th>LHS$_{t+1}$</th>
<th>$\hat{a}_0$</th>
<th>s.e.</th>
<th>$\hat{a}_1$</th>
<th>s.e.</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_t$</td>
<td>$r_{t+1} - g_{t+1}$</td>
<td>$-0.060$</td>
<td>$[0.032]$</td>
<td>$4.847$</td>
<td>$[1.957]$</td>
<td>$6.05%$</td>
</tr>
<tr>
<td></td>
<td>$r_{t+1}$</td>
<td>$-0.000$</td>
<td>$[0.021]$</td>
<td>$3.133$</td>
<td>$[1.157]$</td>
<td>$4.57%$</td>
</tr>
<tr>
<td></td>
<td>$-g_{t+1}$</td>
<td>$-0.060$</td>
<td>$[0.032]$</td>
<td>$1.714$</td>
<td>$[2.080]$</td>
<td>$1.31%$</td>
</tr>
<tr>
<td>$dp_t$</td>
<td>$r_{t+1} - g_{t+1}$</td>
<td>$0.307$</td>
<td>$[0.128]$</td>
<td>$0.068$</td>
<td>$[0.030]$</td>
<td>$4.39%$</td>
</tr>
<tr>
<td></td>
<td>$r_{t+1}$</td>
<td>$0.262$</td>
<td>$[0.079]$</td>
<td>$0.050$</td>
<td>$[0.019]$</td>
<td>$4.29%$</td>
</tr>
<tr>
<td></td>
<td>$-g_{t+1}$</td>
<td>$0.045$</td>
<td>$[0.123]$</td>
<td>$0.018$</td>
<td>$[0.029]$</td>
<td>$0.53%$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RHS(t)</th>
<th>LHS(t+1)</th>
<th>(\hat{a}_0)</th>
<th>s.e.</th>
<th>(\hat{a}_1)</th>
<th>s.e.</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>-0.059</td>
<td>[0.037]</td>
<td>4.711</td>
<td>[2.228]</td>
<td>6.30%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>0.001</td>
<td>[0.021]</td>
<td>3.044</td>
<td>[1.131]</td>
<td>4.47%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>-0.060</td>
<td>[0.042]</td>
<td>1.667</td>
<td>[2.616]</td>
<td>0.82%</td>
</tr>
<tr>
<td>(dp_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>0.301</td>
<td>[0.142]</td>
<td>0.067</td>
<td>[0.033]</td>
<td>4.41%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>0.272</td>
<td>[0.079]</td>
<td>0.053</td>
<td>[0.019]</td>
<td>4.63%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>0.029</td>
<td>[0.154]</td>
<td>0.014</td>
<td>[0.036]</td>
<td>0.21%</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>RHS(t)</th>
<th>LHS(t+1)</th>
<th>(\hat{a}_0)</th>
<th>s.e.</th>
<th>(\hat{a}_1)</th>
<th>s.e.</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>-0.048</td>
<td>[0.038]</td>
<td>2.687</td>
<td>[1.009]</td>
<td>8.44%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>-0.031</td>
<td>[0.052]</td>
<td>3.880</td>
<td>[1.232]</td>
<td>10.32%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>-0.017</td>
<td>[0.043]</td>
<td>-1.193</td>
<td>[1.158]</td>
<td>1.83%</td>
</tr>
<tr>
<td>(dp_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>0.365</td>
<td>[0.116]</td>
<td>0.094</td>
<td>[0.034]</td>
<td>8.58%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>0.565</td>
<td>[0.148]</td>
<td>0.136</td>
<td>[0.045]</td>
<td>10.48%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>-0.200</td>
<td>[0.133]</td>
<td>-0.042</td>
<td>[0.039]</td>
<td>1.85%</td>
</tr>
</tbody>
</table>

Table 5: Predictive regressions for NYSEVW, annual data, 1947–2016.

<table>
<thead>
<tr>
<th>RHS(t)</th>
<th>LHS(t+1)</th>
<th>(\hat{a}_0)</th>
<th>s.e.</th>
<th>(\hat{a}_1)</th>
<th>s.e.</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>-0.033</td>
<td>[0.041]</td>
<td>2.329</td>
<td>[1.087]</td>
<td>7.05%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>-0.013</td>
<td>[0.050]</td>
<td>3.455</td>
<td>[1.172]</td>
<td>9.84%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>-0.020</td>
<td>[0.045]</td>
<td>-1.126</td>
<td>[1.236]</td>
<td>1.85%</td>
</tr>
<tr>
<td>(dp_t)</td>
<td>(r_{t+1} - g_{t+1})</td>
<td>0.315</td>
<td>[0.125]</td>
<td>0.078</td>
<td>[0.036]</td>
<td>7.07%</td>
</tr>
<tr>
<td></td>
<td>(r_{t+1})</td>
<td>0.509</td>
<td>[0.140]</td>
<td>0.117</td>
<td>[0.042]</td>
<td>10.15%</td>
</tr>
<tr>
<td></td>
<td>(-g_{t+1})</td>
<td>-0.194</td>
<td>[0.136]</td>
<td>-0.039</td>
<td>[0.040]</td>
<td>2.01%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RHS&lt;sub&gt;t&lt;/sub&gt;</th>
<th>LHS&lt;sub&gt;t+1&lt;/sub&gt;</th>
<th>( \hat{a}_0 )</th>
<th>s.e.</th>
<th>( \hat{a}_1 )</th>
<th>s.e.</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>-0.077</td>
<td>[0.043]</td>
<td>3.196</td>
<td>[1.162]</td>
<td>9.51%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>-0.027</td>
<td>[0.051]</td>
<td>3.129</td>
<td>[1.243]</td>
<td>5.09%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>-0.050</td>
<td>[0.038]</td>
<td>0.067</td>
<td>[0.946]</td>
<td>0.00%</td>
</tr>
<tr>
<td>( dp_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>0.391</td>
<td>[0.135]</td>
<td>0.104</td>
<td>[0.039]</td>
<td>7.75%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>0.444</td>
<td>[0.158]</td>
<td>0.106</td>
<td>[0.047]</td>
<td>4.46%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>-0.052</td>
<td>[0.117]</td>
<td>-0.002</td>
<td>[0.035]</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table 7: Predictive regressions for NYSEVW, annual data, 1926–2016.

<table>
<thead>
<tr>
<th>RHS&lt;sub&gt;t&lt;/sub&gt;</th>
<th>LHS&lt;sub&gt;t+1&lt;/sub&gt;</th>
<th>( \hat{a}_0 )</th>
<th>s.e.</th>
<th>( \hat{a}_1 )</th>
<th>s.e.</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>-0.066</td>
<td>[0.042]</td>
<td>2.972</td>
<td>[1.135]</td>
<td>8.55%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>-0.011</td>
<td>[0.049]</td>
<td>2.798</td>
<td>[1.153]</td>
<td>4.84%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>-0.055</td>
<td>[0.040]</td>
<td>0.174</td>
<td>[1.010]</td>
<td>0.04%</td>
</tr>
<tr>
<td>( dp_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>0.352</td>
<td>[0.132]</td>
<td>0.091</td>
<td>[0.038]</td>
<td>6.67%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>0.402</td>
<td>[0.144]</td>
<td>0.092</td>
<td>[0.043]</td>
<td>4.34%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>-0.050</td>
<td>[0.121]</td>
<td>-0.001</td>
<td>[0.036]</td>
<td>0.00%</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>RHS&lt;sub&gt;t&lt;/sub&gt;</th>
<th>LHS&lt;sub&gt;t+1&lt;/sub&gt;</th>
<th>( \hat{a}_0 )</th>
<th>s.e.</th>
<th>( \hat{a}_1 )</th>
<th>s.e.</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>-0.140</td>
<td>[0.042]</td>
<td>4.453</td>
<td>[0.980]</td>
<td>12.64%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>0.046</td>
<td>[0.039]</td>
<td>0.928</td>
<td>[0.881]</td>
<td>0.77%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>-0.186</td>
<td>[0.031]</td>
<td>3.525</td>
<td>[0.778]</td>
<td>22.83%</td>
</tr>
<tr>
<td>( dp_t )</td>
<td>( r_{t+1} - g_{t+1} )</td>
<td>0.495</td>
<td>[0.127]</td>
<td>0.138</td>
<td>[0.038]</td>
<td>8.59%</td>
</tr>
<tr>
<td></td>
<td>( r_{t+1} )</td>
<td>0.209</td>
<td>[0.110]</td>
<td>0.038</td>
<td>[0.033]</td>
<td>0.92%</td>
</tr>
<tr>
<td></td>
<td>( -g_{t+1} )</td>
<td>0.286</td>
<td>[0.095]</td>
<td>0.100</td>
<td>[0.028]</td>
<td>12.97%</td>
</tr>
</tbody>
</table>

Recall the linear approximation (19):

\[ r_{t+1} - g_{t+1} = \frac{1}{1-\rho} y_t - \frac{\rho}{1-\rho} y_{t+1}. \]

If \( y_t \) follows an AR(1) with autocorrelation \( \phi \), then this reduces to

\[ E_t (r_{t+1} - g_{t+1}) = \frac{\rho(\phi - 1)}{1-\rho} \bar{y} + \frac{1-\rho\phi}{1-\rho} y_t. \]  \hspace{1cm} (31)

In the body of the paper, we estimate the predictive relationship between \( r_{t+1} - g_{t+1} \) and the predictor variable \( y_t \) (and \( dp_t \)) via linear regression. Under our AR(1) assumption, we could also estimate the constant term and the coefficient on \( y_t \) directly, as in (31), by estimating \( \rho \) and the autocorrelation \( \phi \). Table 11 shows that both approaches give similar results.

<table>
<thead>
<tr>
<th>Method</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>-0.067</td>
<td>3.415</td>
<td>7.73%</td>
</tr>
<tr>
<td>AR(1)</td>
<td>-0.079</td>
<td>3.807</td>
<td>7.63%</td>
</tr>
</tbody>
</table>

Table 11: Comparison of AR(1) parametrization and linear regression. Annual price and dividend data, 1947–2017, from CRSP (cash reinvestment), as in Table 1.
C  Figures using full-sample information

Figure 6: Sentiment indicators calculated using full-sample information to estimate the relationship between $y_t$ (or $dp_t$) and $r_{t+1} - g_{t+1}$.

Figure 7: Bubble indicator vs. detrended log volume (full sample)
D Mathematics

D.1 Derivation of (30)

According to equation (10) in result 2 and if now dividend growth is completely unforecastable

\[ y_t = (1 - \rho) \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t [r_{t+1+i} - g_{t+1+i}] \]

\[ = (1 - \rho) \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t [r_{t+1+i}] - \bar{g} \]

\[ = (1 - \rho) \mathbb{E}_t [r_{t+1}] + (1 - \rho) \rho \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t [r_{t+2+i}] - \bar{g}. \]

The above implies

\[ \frac{\mathbb{E}_t [r_{t+1}] - y_t - \bar{g}}{\rho} = \mathbb{E}_t [r_{t+1}] - (1 - \rho) \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t [r_{t+2+i}]. \]

Using the following inequality

\[ \mathbb{E}_t [r_{t+1}] - y_t - \bar{g} \geq \tilde{B}_t - \bar{g} = \text{LVIX}_t + r_{f,t} - y_t - \bar{g}, \]

we finally arrive at

\[ \mathbb{E}_t [r_{t+1}] - (1 - \rho) \sum_{i=0}^{\infty} \rho^i \mathbb{E}_t [r_{t+2+i}] \geq \frac{\tilde{B}_t - \bar{g}}{\rho} \]

as in equation (30).