

JOURNAL OF Financial ECONOMICS

Journal of Financial Economics 44 (1997) 169–203

Book-to-market, dividend yield, and expected
market returns: A time-series analysis

S.P. Kothari, Jay Shanken*

*William E. Simon Graduate School of Business Administration, University of Rochester,
Rochester, NY 14627, USA*



Book-to-market, dividend yield, and expected market returns: A time-series analysis

S.P. Kothari, Jay Shanken*

*William E. Simon Graduate School of Business Administration, University of Rochester,
Rochester, NY 14627, USA*

Received August 1995; received in revised form August 1996

Abstract

We find reliable evidence that both book-to-market (B/M) and dividend yield track time-series variation in expected real stock returns over the period 1926–91 (in which B/M is stronger) and the subperiod 1941–91 (in which dividend yield is stronger). A Bayesian bootstrap procedure implies that an investor with prior belief 0.5 that expected returns on the equal-weighted index are never negative comes away from the full-period B/M evidence with posterior probability 0.08 for the hypothesis (0.14 with the impact of the 1933 outlier tempered). Although this raises doubts about market efficiency, the post-1940 evidence is consistent with expected returns always being positive.

Keywords: Book-to-market; Dividend yield; Expected returns; Bootstrap; Bayesian
JEL classification: G12; G14; C11; C15

1. Introduction

There is considerable evidence that stock returns are predictable. The academic literature on return predictability goes back at least to the late 1970s.

*Corresponding author.

We thank Ray Ball, Peter Bossaerts, Steve Cauley, Gene Fama, Ken French (the referee), Burton Hollifield, John Long, Bill Schwert (the editor), Rob Stambaugh, Jerry Warner, Robert Whitelaw, and especially Johnathan Lewellen. Also, thanks to participants in seminars at NBER, the Southern Finance Association, Virginia Tech, London Business School, Dartmouth, Washington University, Yale, the University of California at Irvine and Los Angeles, Cambridge, and University of Rochester. We gratefully acknowledge financial support from the John M. Olin Foundation and the Bradley Policy Research Center at the Simon School, University of Rochester. Assistance from Kim Goetz and Jane Muellner on manuscript preparation is appreciated.

Several studies (e.g., Fama and Schwert, 1977; Keim and Stambaugh, 1986; and Fama and French, 1988) regress returns on predetermined variables, including past returns, to infer the existence of statistically significant time-variation in expected returns. Among the predetermined variables, many studies have found dividend yield to have predictive power in both cross-section (Litzenberger and Ramaswamy, 1979) and time series (Rozeff, 1984; Shiller, 1984; Fama and French, 1988a, 1989; and Campbell and Shiller, 1988). Since the publication of Fama and French (1992, 1993), the book-to-market ratio (B/M) has emerged as a strong contender as a determinant of expected returns. Fama and French, Davis (1994), and Chan et al. (1995), among others, provide evidence that B/M significantly explains cross-sectional variation in average returns, although Kothari et al. (1995) show that the effect is weaker in large firms (also see Loughran, 1996) and argue that the magnitude and significance of the effect may be overstated due to data mining and a variety of selection biases in the COMPUSTAT data base (also see Breen and Korajczyk, 1994).

The literature emphasizes two views of the predictive power of financial ratios that are relevant in both time series and cross-section. The first (efficient markets) view has been referred to by Fama and French (1988b) as the 'discount-rate' effect (also see Keim and Stambaugh, 1986). Holding expected cash flows roughly constant, anything that increases the rate at which cash flows are discounted (e.g., changes in risk or liquidity) lowers market value and increases both expected return and financial ratios. This assumes, of course, that the most recent value in the numerator (dividend or book value of equity) does an adequate job of controlling for expected future cash flow. An alternative (inefficient markets) view is that financial ratios reflect the extent to which the market is overvalued (low ratios) or undervalued (high ratios) at a given point in time. In the case of underpricing, for example, future returns (and thus 'true' expected returns) will be high insofar as the undervaluation is likely to be corrected over the given horizon.¹

In this research we evaluate the ability of an aggregate B/M ratio to track time-series variation in expected market index returns, and compare its forecasting ability to that of dividend yield. We employ a vector-autoregressive (VAR) framework to evaluate the statistical significance of the regression evidence of predictability of expected returns. The bootstrap simulation procedure is used to test the null hypothesis of no predictability as well as to assess the magnitude and economic significance of the relation between expected returns and its determinants, such as B/M and dividend yield. Finally, we extend the traditional bootstrap procedure by developing a Bayesian-bootstrap simulation. With the

¹ A third motivation for the cross-sectional investigation of dividend yield is taxes. This study, however, does not focus on the tax motivation.

exceptions of MacBeth and Emanuel (1993) and independent work by Pontiff and Schall (1995), B/M has not been considered in time series and the relative abilities of B/M and dividend yield have not been evaluated.²

Our time-series investigation at the market level employs the Center for Research in Security Prices (CRSP) equal-weighted and value-weighted portfolio annual returns data. Real returns are computed using the observed CPI inflation rate. Corresponding yields equal to dividends paid in the preceding year divided by the price level at the end of that year are also computed from CRSP data as in Fama and French (1988a). Time-series data on B/M are for the stocks included in calculating the Dow Jones Industrial Average. Return and B/M data are available since 1926, permitting us to study B/M over a longer period than has been possible in the cross-sectional literature. Use of the Dow B/M variable along with broader index returns is in the spirit of Fama and French (1989), who regress bond returns as well as stock index returns on both stock yields and bond-yield spreads, viewing the explanatory variables as proxies for the overall state of the aggregate economy.

Although the use of ordinary least squares (OLS) regressions to assess the ability of dividend yield and B/M to forecast one-year returns might appear to be straightforward, time-series regressions of returns on beginning-of-period yield measures are subject to small-sample biases (Stambaugh, 1986; Mankiw and Shapiro, 1986). The bias arises from the fact that the regression disturbances are correlated with future values of the independent variable. A large increase in price tends to produce both large positive returns and a contemporaneous decrease in yield. Since the yield only gradually reverts to a normal level, large returns tend to follow what, in sample, appear to be relatively high yields. This imparts an upward bias to the regression slope coefficient. The bias is opposite to that which arises when the lagged dependent variable is included as a regressor, since the beginning-of-period yield is highly negatively correlated with the lagged return. In large samples, the sample mean of the yield measure converges to the true mean (under fairly general assumptions), eliminating the source of the spurious correlation between yield and return.

² MacBeth and Emanuel (1993) presents evidence that price-to-book has some forecasting ability for annualized 10-year returns on the Standard and Poor's Industrial Index from 1946 to 1982. It is not clear (to us) from the discussion in the paper, however, whether the price and book values simulated under the null hypothesis were related in a manner similar to that observed in the historical data. If not, there is some question as to the reliability of the results. We have similar concerns about the bootstrap procedure of Goetzmann and Jorion (1993), in which there is no relation between randomized returns and the (actual) dividends used in computing dividend yield. It is natural, under the null hypothesis, to have returns independent of past dividends, but returns should anticipate future dividends in an efficient market, even if expected returns are constant (e.g., see the evidence in Kothari and Shanken, 1992). The VAR framework discussed below allows this feature of the data to be reflected in the simulated model.

Since large-sample inference is suspect in this context, we resort to simulation methods to evaluate the statistical significance of the regression evidence on return predictability, as in recent studies by Hodrick (1992), Goetzmann and Jorion (1993), and Nelson and Kim (1993). In particular, we employ a VAR framework to simulate an empirical distribution for the OLS slope coefficient. As previous authors have noted, the VAR simulation approach is appealing in that it explicitly models the process relating regression disturbances to future values of the independent variable. Moreover, the complications introduced by using overlapping multiyear returns (the usual standard errors are too low) can be addressed, along with conditional heteroskedasticity of the disturbances. The existing simulation evidence indicates that conventional asymptotic methods are not reliable; even in the nonoverlapping return applications considered here.

Thus far, the simulation literature has focused on testing the point null hypothesis of no predictability in expected returns, i.e., a slope coefficient of zero in the return regression. While this is a natural place to start, it would be useful to be able to say something more about the magnitude and economic significance of the true coefficient if the null hypothesis is rejected. We explore the issue of economic significance by evaluating the power function of the test in simulations. This permits a formal test of the composite null hypothesis that the slope coefficient is less than some prespecified value. Fama and French (1988a) provide some indication of economic significance and the impact of bias by looking at the out-of-sample forecasting performance of their estimated return/yield relation. Degrees of freedom are a scarce resource when using multi-period returns, however, limiting the usefulness of this approach. Hodrick (1992) also simulates a particular alternative to explore the power of his tests.

In addition, we develop a Bayesian-bootstrap simulation procedure that estimates the *likelihood* of different slope values, conditional on the historical OLS estimate. The information contained in the likelihood function nicely complements that conveyed by conventional *p*-value analysis and provides a richer framework in which to assess economic significance. Moreover, the Bayesian interpretation of the likelihood function allows us to easily incorporate prior beliefs about plausible expected return variation and to explore the impact of different views of market efficiency on the conclusions drawn from the given sample evidence.

Overall, we find evidence of economically and statistically significant variation in one-year returns forecast by the B/M ratio over the 1926–91 period. The slope coefficients, especially for the equal-weighted index, are so large as to suggest that expected real returns are sometimes negative, raising doubts about market efficiency. The evidence for dividend yield is comparable for the value-weighted index returns and weaker for the equal-weighted index. In the 1941–91 subperiod, both B/M and dividend yield track expected return variation and the evidence is consistent with positive expected returns. The B/M and dividend

yield results are comparable for the equal-weighted index, but the dividend yield evidence is stronger for the value-weighted index in the subperiod.

We report descriptive statistics in Section 2 and initial regression results in Section 3. Section 4 describes our bootstrap simulation methodology, while Section 5 describes the computation of bootstrap *p*-values and likelihoods based on the simulations. Section 6 discusses the role of prior information about expected returns in the context of testing the sensitivity of expected returns to determinants like B/M and dividend yield. Bootstrap simulation results are presented in Sections 7 and 8. Section 7 focuses on the magnitude of the slope coefficients on the financial ratios in the expected return relation, while Section 8 provides tests of the hypothesis that expected returns are always nonnegative. Section 9 summarizes our main findings.

2. Descriptive statistics

This section begins with a brief description of the data used, followed by descriptive statistics for the time series of B/M, dividend yield, and market index returns. To ensure that the B/M ratio is known to investors at the beginning of the return year, we measure annual returns from April of year t to March of year $t + 1$ and regress these returns on the ratio of book value for year $t - 1$ to price at the end of March of year t . The corresponding dividend yield equals dividends paid from April of year $t - 1$ to March of year t divided by price at the end of March of year t . These ratios and returns are referred to as the 'year t ' observations.

The book value 'per share' that corresponds to the Dow Jones Industrial Average is published in *The Value Line Investment Survey*. Time series data are available since 1920. Book value data for 1920–35 are based on estimates by *The Value Line*. The book value for year $t - 1$ is deflated by the Dow Jones Industrial Average at the end of March of year $t + 1$ to obtain the B/M ratio, which is used to forecast returns for year t .

We begin by presenting some descriptive statistics in Table 1 for returns, B/M, and dividend yield over the sample period 1926–91 and subperiod 1941–91. Return and dividend yield information is for both equal- and value-weighted CRSP indexes. All three series are positively skewed, with the equal-weighted returns being the most highly skewed. For the whole period, the average real return on the equal-weighted (value-weighted) index is 15.3% (9.4%) with standard deviation 38.8% (23.9%). The minimum is -61% (-48.4%) in 1937 and the maximum is 214% (85.7%) in 1933. The returns on both indexes are lower and less volatile in the post-1940 subperiod.

The mean of the B/M series for 1926–91 is 70% with standard deviation 23%. The minimum is 27% for 1929 and the maximum is 148% for 1933, the same year in which returns achieve their maximum for both indexes. The mean of the

Table 1
Descriptive statistics for the return and financial ratio variables

Variable	Mean	Std. dev.	Skewness	Min	Q1	Median	Q3	Max
<i>Panel A: Period 1926–91</i>								
Equal-weighted return (%)	15.3	38.8	2.30	–60.9	–6.3	10.2	27.7	213.9
Value-weighted return (%)	9.4	23.9	0.36	–48.4	–4.3	9.0	21.4	85.7
Book-to-market (%)	69.5	22.7	0.87	27.2	52.8	66.5	80.5	147.7
Equal-weighted dividend yield (%)	3.65	1.44	1.09	1.41	2.71	3.22	4.31	8.21
Value-weighted dividend yield (%)	4.40	1.42	1.41	2.66	3.33	4.11	5.06	9.27
<i>Panel B: Period 1941–91</i>								
Equal-weighted return (%)	12.5	23.2	0.53	–31.0	–3.0	9.4	25.0	75.2
Value-weighted return (%)	8.6	17.2	0.07	–27.1	–3.7	8.4	19.7	42.2
Book-to-market (%)	70.7	19.0	0.59	42.8	54.7	69.0	81.4	118.6
Equal-weighted dividend yield (%)	3.67	1.35	0.94	1.85	2.71	3.18	4.51	7.27
Value-weighted dividend yield (%)	4.24	1.23	1.04	2.66	3.21	4.05	5.06	8.39

Equal-weighted return is the CRSP equal-weighted annual real return inclusive of dividends from April of year t to March of year $t + 1$.

Value-weighted return is the CRSP value-weighted annual real return inclusive of dividends from April of year t to March of year $t + 1$.

Book-to-market is the ratio of the book value per share at the end of year $t - 1$ of the stocks in the Dow Jones Industrial Average to the Dow Jones Industrial Average at the end of March of year t .

Equal-weighted dividend yield is the ratio of dividends paid from April of year $t - 1$ to March of year t to the accumulated value of one dollar invested in the CRSP equal-weighted index exclusive of dividends for one year ending on March 31 of year t .

Value-weighted dividend yield is the ratio of dividends paid from April of year $t - 1$ to March of year t to the accumulated value of one dollar invested in the CRSP value-weighted index exclusive of dividends for one year ending on March 31 of year t .

equal-weighted dividend yield series is 3.65% with standard deviation 1.44%. The minimum is 1.41% for 1934 and the maximum is 8.21% for 1938. Dividend yields for the value-weighted index are higher with a mean of 4.40%. All ratios are less volatile over the post-1940 subperiod.

Both financial ratios are highly autocorrelated, with 1926–91 first-order autocorrelations of 0.68 for B/M and 0.56 (0.55) for the equal-weighted (value-weighted) index dividend yield. We cannot reject the null hypothesis that each series is generated by a first-order autoregressive process. The correlation between B/M and equal-weighted (value-weighted) dividend yield is 0.42 (0.66). The higher correlation with the value-weighted yield is reasonable, since the B/M ratio is for the relatively large firms in the Dow index. The financial ratio autocorrelations are all higher in the post-1940 period, close to 0.80. Also, B/M is 10–15% more highly correlated with each dividend yield series.

3. OLS regression results

This section summarizes OLS regression results, deferring an assessment of small-sample biases to later sections. Table 2 presents the results of regressing annual returns on B/M (panel A) or dividend yield (panel B). The table reports results using both equal- and value-weighted index returns for the entire sample period 1926–91. We also explore the sensitivity of results to the 1933 outlier by running the regressions with the outlier deleted or ‘truncated’. In the latter case, the regressions are run as if the extreme 1933 values for return (214% equal-weighted, 86% value-weighted) and B/M (148%) had been the next highest sample values (129% equal-weighted return, 74% value-weighted return, and 119% B/M). Since systematic deletion of the single observation most supportive of a positive regression slope is obviously biased against such a finding, truncation is adopted as a more neutral means of evaluating sensitivity. This approach is in the spirit of the statistical technique known as Winsorization (see Dixon and Massey, 1969).

The standard errors and t -statistics (except White t) are the usual OLS quantities. The B/M and dividend yield variables are scaled by their respective standard deviations so that the slopes can be interpreted as the change in expected return for a one-standard deviation change in the values of the financial ratios. This helps in assessing the economic significance of a given beta and facilitates the formulation of one’s prior beliefs. The full-sample (‘All’) slope estimates in Table 2 are all large in economic terms and most are more than two or three standard errors above zero in these simple regressions.

The 1926–91 slope coefficient on B/M using the equal-weighted index is 20.2%, with OLS standard error 4.1% and t -statistic 4.87 (p -value 0.000). The adjusted R^2 of nearly 26% is quite large relative to past findings for 1-year returns, and the estimated slope is enormous. If B/M increases from its historical

Table 2
Univariate OLS regression results: 1926–91

Index	Sample	a in %	$s(a)$ in %	t -stat	b in %	$s(b)$ in %	t -stat	White t -stat	Adjusted R^2 in %
Panel A: $R_{t+1} = \alpha + \beta B/M_t + u_{t+1}$									
Equal weight	All	-46.7	13.4	-3.49	20.2	4.1	4.87	2.68	25.9
	1933 deleted	-23.0	12.3	-1.88	10.6	3.5	3.00	3.33	11.1
	1933 truncated	-32.1	12.6	-2.54	14.3	3.7	3.81	3.37	17.2
Value weight	All	-16.9	9.0	-1.87	8.6	2.8	3.07	2.48	11.5
	1933 deleted	-8.3	9.4	-0.88	5.0	2.7	1.84	1.90	3.6
	1933 truncated	-13.6	9.4	-1.45	7.1	2.8	2.54	2.31	7.7
Panel B: $R_{t+1} = \alpha + \beta DYId_t + u_{t+1}$									
Equal weight	All	-1.7	13.0	-0.13	6.7	4.8	1.39	2.02	1.4
	1933 deleted	-5.9	10.0	-0.59	7.2	3.7	1.95	2.18	4.2
	1933 truncated	-3.44	11.0	-0.31	6.9	4.1	1.69	2.09	2.8
Value weight	All	-12.7	9.3	-1.37	7.1	2.9	2.50	2.26	7.5
	1933 deleted	-8.0	8.9	-0.90	5.2	2.7	1.92	2.01	4.0
	1933 truncated	-11.9	9.1	-1.31	6.8	2.8	2.44	2.29	7.1

R_{t+1} is the CRSP equal-weighted or value-weighted real return inclusive of dividends from April of year t to March of year $t+1$. B/M_t is the ratio of the book value per share at the end of year $t-1$ of the stocks in the Dow Jones Industrial Average to the Dow Jones Industrial Average at the end of March of year t . B/M_t is deflated by its standard deviation over the sample period. $DYId_t$ is the ratio of dividends paid from April of year $t-1$ to March of year t to the accumulated value of one dollar invested in the CRSP equal- or value-weighted index exclusive of dividends for one year ending on March 31 of year t . $DYId_t$ is deflated by its standard deviation over the sample period.

average of 70% by one standard deviation (23%) the expected real return, which averages 15.3%, is forecast to rise by 20.2 percentage points. As we shall see later, adjustments for small-sample bias do not materially alter these conclusions. Deleting the 1933 outlier cuts the B/M coefficient in half, but the estimate is still very large and highly statistically significant. Truncation also has a substantial, but smaller, effect. Turning to the regression results for the value-weighted index reveals that the Dow B/M ratio explains a much smaller fraction of the time-series variation in these returns than for the CRSP equal-weighted index.

Dividend yield, in contrast, is somewhat better at explaining variation in returns on the value-weighted index than on the equal-weighted index. For the full sample, dividend yield explains less of the variation in value-weighted index returns than does B/M, but explanatory power is similar with deletion or truncation. Although equal-weighted dividend yield steadily increases from 1929 to 1932, it drops nearly by half, to 3.44%, at the beginning of 1933. Hence, deleting the 1933 observation actually strengthens the relation between equal-weighted returns and dividend yield. Although value-weighted dividend yield also begins to drop, at 6.94% it is still well above its mean 4.40% at the beginning of 1933 and hence deletion weakens the value-weighted relation.

The slightly better ability of dividend yield to explain value-weighted index returns may be related to the fact that a large fraction of the firms in the index do not pay dividends. Therefore, these firms' dividend yields cannot reflect changes in expected market return. Since firms that do not pay dividends are, on average, smaller in market capitalization, they have a relatively small effect on the dividend yield of the value-weighted index.

Multiple regressions of returns on B/M and dividend yield over the 1926–91 period, shown in Table 3, tell essentially the same story as the univariate regressions. B/M dominates dividend yield in explaining equal-weighted index returns. B/M also outperforms dividend yield for value-weighted index returns when the 1933 outlier is included, but otherwise the explanatory power of the variables is similar.

Given the concerns about small-sample biases mentioned earlier, we can hardly have much confidence in the White (1980) asymptotic adjustment for heteroskedasticity shown in Table 2. We use it as a heuristic for (we hope) identifying situations in which results may be particularly sensitive to heteroskedasticity problems or other misspecifications. This appears to be the case for the full-sample equal-weighted index regressions, in which the White t -statistic is about 45% smaller (larger) than the usual OLS t -statistic for the B/M (dividend yield) slope. Exploratory regressions (not reported) of squared or absolute return residuals on B/M suggest a significantly positive conditional variance relation, whereas the relation for dividend yield is negative and statistically insignificant. We examine weighted least squares and nonlinear regression specifications for B/M, with little impact on our basic conclusions. There is no

Table 3

OLS multiple regression results: 1926–91

$$R_{t+1} = \alpha + \beta_1 B/M_t + \beta_2 DYld_t + u_{t+1}$$

Index	Sample	a (%)	$s(a)$ (%)	t -stat	b_1 (%)	$s(b_1)$ (%)	t -stat	White t -stat	b_2 (%)	$s(b_2)$ (%)	t -stat	White t -stat	Adj. R^2 in %
Equal weight	All	-43.8	14.6	-3.00	21.1	4.6	4.60	2.32	-2.3	4.6	-0.49	-0.41	25.0
	1933 deleted	-25.6	12.9	-1.99	9.3	4.1	2.30	2.25	2.7	4.1	0.68	0.65	10.3
	1933 truncated	-32.7	13.5	-2.41	14.0	4.2	3.32	2.58	0.5	4.2	0.13	0.12	15.9
Value weight	All	-19.4	9.8	-1.97	6.9	3.8	1.81	1.38	2.5	3.8	0.65	0.59	10.7
	1933 deleted	-11.4	10.0	-1.14	2.8	3.6	0.76	0.67	3.4	3.6	0.93	0.83	3.4
	1933-truncated	-17.0	10.0	-1.70	4.5	3.8	1.19	1.03	3.7	3.8	0.98	0.89	7.7

R_{t+1} is the CRSP equal-weighted or value-weighted real return inclusive of dividends from April of year t to March of year $t+1$.

B/M_t is the ratio of the book value per share at the end of year $t-1$ of the stocks in the Dow Jones Industrial Average to the Dow Jones Industrial Average at the end of March of year t . B/M_t is deflated by its standard deviation over the sample period.

$DYld_t$ is the ratio of dividends paid from April of year $t-1$ to March of year t to the accumulated value of one dollar invested in the CRSP equal- or value-weighted index exclusive of dividends for one year ending on March 31 of year t . $DYld_t$ is deflated by its standard deviation over the sample period.

Table 4

OLS regression results: 1941–91

Index	a in %	$s(a)$ in %	t -stat	b in %	$s(b)$ in %	t -stat	White t -stat	Adjusted R^2 in %
Panel A: $R_{t+1} = \alpha + \beta B/M_t + u_{t+1}$								
Equal weight	-19.9	11.9	-1.68	8.7	3.1	2.83	2.79	12.3
Value weight	-7.8	9.2	-0.85	4.4	2.4	1.86	2.01	4.7
Panel B: $R_{t+1} = \alpha + \beta DYld_t + u_{t+1}$								
Equal weight	-2.7	8.8	-1.45	9.3	3.0	3.06	2.80	14.3
Value weight	-15.0	8.1	-1.86	6.9	2.3	3.05	3.20	14.3

$$Panel C: R_{t+1} = \alpha + \beta_1 B/M_t + \beta_2 DYld_t + u_{t+1}$$

Index	a	$s(a)$	t -stat	b_1	$s(b_1)$	t -stat	White t -stat	b_2	$s(b_2)$	t -stat	White t -stat	Adj. R^2 in %
Equal weight	-25.7	11.9	-2.15	5.4	3.4	1.58	1.68	6.60	3.44	1.92	1.87	16.8
Value weight	-13.8	9.1	-1.51	-1.0	3.2	-0.31	-0.31	7.57	3.24	2.34	2.28	12.7

R_{t+1} is the CRSP equal-weighted or value-weighted real return inclusive of dividends from April of year t to March of year $t+1$.

B/M_t is the ratio of the book value per share at the end of year $t-1$ of the stocks in the Dow Jones Industrial Average to the Dow Jones Industrial Average at the end of March of year t . B/M_t is deflated by its standard deviation over the sample period.

$DYld_t$ is the ratio of dividends paid from April of year $t-1$ to March of year t to the accumulated value of one dollar invested in the CRSP equal- or value-weighted index exclusive of dividends for one year ending on March 31 of year t . $DYld_t$ is deflated by its standard deviation over the sample period.

significant association between B/M and squared or absolute residuals (t -statistics less than 0.5) with the 1933 outlier deleted, and both methods of treating the outlier greatly reduce the huge gap between conventional and White t -statistics.

Table 4 contains regression results for the 1941–91 subperiod. The motivation for examining this subperiod is that, as noted by Officer (1973) and Schwert (1989), stock return variability was unusually high during the 1929–39 Great Depression. Dividend yield modestly outperforms B/M over the subperiod with equal-weighted index returns and substantially outperforms B/M with value-weighted returns. Analysis of post-1962 data reveals that dividend yield tracks variation in expected returns through time, while the evidence for B/M is inconclusive. Our full-period and subperiod results are similar to those reported by Pontiff and Schall (1995), who also include default and term premiums and three-month T-bill yield in their regression model and whose subperiod begins in 1959.

Overall, we find that both B/M and dividend yield appear to track variation in market index returns, but full- and subperiod analyses suggest that neither variable consistently dominates the other in its explanatory power. Apart from the dramatic full-period, equal-weighted index (B/M case), the OLS slope estimates range from 4.4% to 9.3%, still large in economic terms. These values of beta imply that a one-standard-deviation change in the value of B/M or dividend yield is associated with a shift of about 4–9% in the expected real return on the index.

4. The simulation procedure

We now consider simulation methods to examine whether the conclusions above are materially affected by small-sample biases. Like Nelson and Kim (1993), we posit a simple VAR model of returns and determinants of expected returns. In this model, the expected value at time t of the return realized at $t + 1$, r_{t+1} , is a linear function of the variable x_t (dividend yield or B/M), which follows a simple autoregressive process:

$$r_{t+1} = \alpha + \beta x_t + u_{t+1}; u \sim \text{i.i.d. } (0, \sigma_u^2), \quad (1)$$

$$x_{t+1} = c + \phi x_t + v_{t+1}; v \sim \text{i.i.d. } (0, \sigma_v^2), \quad (2)$$

where u and v are white noise error terms with contemporaneous covariance σ_{uv} . The parameters in Eqs. (1) and (2) are estimated for annual real returns data by OLS.

Stambaugh (1986) shows that the bias in the OLS estimate b of β in Eq. (1) is

$$E(b - \beta) = (\sigma_{uv}/\sigma_v^2)E(p - \phi) \quad (3)$$

where p is the OLS estimate of ϕ in Eq. (2). As expected, the residuals in the return regression equation are strongly negatively contemporaneously correlated with the residuals in the financial ratio autoregression (2). For the 1926–91 period, the correlation is -0.80 in the case of B/M and -0.65 (-0.79) for equal-weighted (value-weighted) dividend yield. Over the post-1940 period, residuals in the two dividend yield VAR's are more strongly contemporaneously correlated, about -0.85 in each case. The negative correlation is induced by the fact that price increases tend to reduce the financial ratios while resulting in positive returns. Thus, by (3), the OLS slope estimates are biased upwards.

Combining (3) with the fact that the bias in p is approximately $-(1 + 3\phi)$ divided by the sample size T (Kendall, 1954) suggests the following bias-adjusted estimator of β :

$$b_A = b + (s_{uv}/s_v^2)(1 + 3p_A)/T, \quad (4)$$

where s_{uv} and s_v^2 are sample moments of the OLS residuals from Eqs. (1) and (2) and $p_A = (Tp + 1)/(T - 3)$ is a bias-adjusted estimator for ϕ . After computing the bias-adjusted estimates, b_A and p_A , the intercepts are specified by setting the average fitted values in (1) and (2) equal to the sample means. Corresponding adjusted (u, v) residual pairs are then obtained from (1) and (2), with each residual series averaging to zero over time. These residuals are used in the bootstrap simulations which we describe below.

The system is started up with the historical value for x at the beginning of the sample period. Given a hypothetical value of β and substituting the estimator p_A for ϕ , time series of returns and new values of x are then generated from (1) and (2) by randomly selecting adjusted (u, v) residual pairs with replacement. Finally, the simulated time series of returns, which has the same length as the original data used to estimate Eq. (1), is regressed on the simulated time series of x 's. The estimated slope from this regression is saved. This procedure is repeated 2500 times with the hypothetical β value fixed and, each time, starting the system with the historical value for x at the beginning of the sample period. This yields an empirical distribution of slope estimates that is employed in a variety of inference procedures.³

The entire process (2500 iterations) is repeated for each of a discrete set of beta values starting from zero. Thus, we have a different empirical distribution of slope estimates for each hypothetical alternative. We focus on nonnegative slope

³ Since the distribution of the residuals is fixed across simulations, as the true slope coefficient varies, the simulated variance of returns increases with the predictable variation in returns. If anything, this would appear to make it more difficult to find evidence of strong predictability and thus might impart a conservative bias to our results. Likewise, with regard to the use of bias-adjusted residuals. It would be interesting to explore the sensitivity of inferences to alternative procedures which attempt to maintain a fixed total variability of returns.

values since the two prominent competing hypotheses concerning predictability, which are market inefficiency (temporary mispricing) and time-variation in equilibrium expected returns (and the associated discount-rate effect on market value), both suggest a positive relation between returns and the financial ratios. The set of beta values considered allows for the possibility that expected returns are negative when the market is overvalued; the maximum beta is specified so as to generate a minimum expected return equal to the negative of the sample mean return. This specification of the domain of the prior distribution is not 'pure' in that it makes some use of the realized data. Obviously, other ranges could be considered, but some heuristic is needed to limit the analysis and this one strikes us as reasonable.

5. Bootstrap p -values and likelihoods

To test the null hypothesis that β equals zero, we look at the proportion of the 2500 simulated slope estimates, generated under the $\beta = 0$ hypothesis, that exceed the historical slope. The latter is the estimate obtained from the actual data. This gives us an empirical p -value for the historical slope. The procedure is similar to that employed by Nelson and Kim (1993), except for our use of bias-adjusted estimates and residuals, and two other differences: they randomly select an initial start up value from the unconditional distribution of x , rather than using the initial historical value (our approach is motivated by the conditionality principle of inference (Berger, 1985, Section 1.6)), and they sample without replacement while we sample with replacement.

If the null hypothesis $\beta = 0$ is rejected, it is of further interest to see whether values of β up to some given positive level can be rejected as well. To test this composite null hypothesis, we look at the empirical distribution of estimates generated under the assumption that β equals the prespecified level. Again, we compute the proportion of estimates that exceed the historical β estimate and report this proportion as an empirical p -value for the composite hypothesis. Since larger slope estimates should be generated when the true slope is larger, we expect to see the empirical p -values increase (weaker rejection) as the prespecified level of β increases. Note that the p -value for testing the composite null hypothesis that b is less than or equal to a given value is just the power function evaluated at that value of beta, for a test with size (type I error probability) equal to the p -value for $\beta = 0$. Shanken (1987a) develops a similar approach to testing a composite hypothesis of approximate portfolio efficiency.

It is well known (but sometimes forgotten) that the p -value is not the probability that the given null hypothesis is true (see Lindley, 1957). Rather, it is just one 'metric' for assessing the extent to which the sample evidence looks unusual under the null hypothesis. Specifically, it is the probability under the null of obtaining a historical slope as big as (or bigger than) that observed. The

distribution of the sample statistic under the alternative plays no role in this context.

In order to obtain probabilities for the different hypotheses, conditional on the historical evidence, the Bayesian approach focuses on the 'likelihood function', in conjunction with prior (subjective) probabilities for the hypotheses. With a continuous density function for the sample evidence (here, the historical slope), the relative likelihood of one hypothesis as compared to another can loosely be described as follows: it is the ratio of the respective probabilities (under the different hypotheses or models) of generating sample evidence within an infinitesimal neighborhood of the historical slope.

If the VAR disturbances in Eqs. (1) and (2) are assumed to be jointly normally distributed, a closed-form expression for the likelihood function can be obtained (see Kandel and Stambaugh, 1996). Here, we take a different approach, introducing what we call a bootstrap likelihood, i.e., the proportion of simulated samples (under a given hypothesis) for which the observed slope is in some fixed neighborhood of the historical slope. Although lacking the mathematical elegance of more formal Bayesian approaches, our strategy, like the 'classical' bootstrap, is simple and pragmatic, and does not require much in the way of assumptions.⁴ For example, the bootstrap procedure allows for the positive skewness that we sometimes observe in the VAR residuals. A similar approach could be adapted to a variety of situations that are commonly viewed as analytically intractable and, therefore, not amenable to Bayesian analysis. While there is a literature on nonparametric density estimation, we are not aware of any papers that have used bootstrapping techniques in the manner of this paper.

Bootstrap likelihoods are computed for the same set of β values considered in the classical p -value analysis. These proportions are scaled so that the sum of the likelihoods over all values of β considered is one. As in any Bayesian application, scaling does not affect the relative likelihoods or the associated posterior probabilities derived from the likelihoods and a given prior. Moreover, the scaled proportions can be interpreted directly as posterior probabilities for the different slope values when prior beliefs are uniformly distributed over the various slopes. Although we do not claim that this is the 'right' prior, it provides a convenient reference point for interpreting the data. Additionally, if the likelihood is (approximately) zero for the largest slope considered, posterior probabilities would be unaffected by expanding the uniform prior to larger values of beta. This is the case in many of our simulations.

⁴ One should not push this point too far, as any bootstrap technique is limited by the informativeness of the observed sample from which the simulations are developed. Also, in principle we would want to condition on a complete set of sufficient statistics for the sample. Thus, our simplified procedure may entail some loss of information, although it is easily extended to accommodate conditioning on several statistics.

We have explored the impact of varying the width of the neighborhood of the historical slope used in computing the likelihoods. The sensitivity to this alteration is minimal (usually identical to two decimals). This robustness is reassuring for our approach and presumably reflects the fact that the true density function is approximately linear within the small neighborhoods considered.

6. Prior information and expected returns

In this section we consider the role of prior information about expected returns and the implications of this information for hypotheses about β . In our linear regression framework, the implied expected return or conditional mean for a given level of the financial ratio, x , depends on both regression parameters, α and β . Intuitively, we tend to think about priors on these parameters by starting with prior beliefs about the general level of expected returns and x , as well as the extent of variation in expected returns for given movements in x . The latter is our prior on β . The prior for α is then implied by these more fundamental beliefs about the level and variability of expected returns.

In some Bayesian applications, researchers impose so-called ‘diffuse’ or ‘non-informative’ priors intended to represent initial ignorance about the parameters of a model; See Section 3.3 of Berger (1985) for a general introduction and DeGroot (1982) for an insightful critique of the use of diffuse priors. This attempt to achieve ‘objectivity’ avoids the often difficult, but rewarding, activity of assessing the economic (or other) meaning of alternative parameter values and structuring the prior beliefs accordingly. Here, we take the view that the prior distribution is a valuable tool for exploring the economic significance of the likelihood function and that it is important to consider a range of informative prior beliefs rather than some abstract notion of the ‘right’ prior. While this will entail some ad hoc assumptions, we believe that the insights obtained justify the structure adopted. Shanken (1987b) entertains a similar Bayesian approach to testing portfolio efficiency. The limitations of diffuse priors in this context are clarified in Kandel et al. (1995).⁵

Recall that α is specified as $\bar{r} - \beta\bar{x}$ in our simulations. Given (1), this entails an assumption that the *sample* mean, \bar{u} , of the return surprises is zero. This is a natural simplifying assumption when the focus is on the slope estimator and its sampling distribution. The condition has important consequences for analysis of the *level* of expected returns, however. As we now argue, the assumption about \bar{u} amounts to an indirect method of characterizing beliefs about α .

⁵ See Markowitz and Usmen (1996) for additional discussion of these issues in another financial context.

Given a value of β and the sample observations \bar{x} and \bar{r} , the condition $\bar{u} = 0$ amounts to an assertion that one’s initial belief about α is consistent with β and the relative values of the mean outcomes. In contrast, since $\bar{r} = \alpha + \beta\bar{x} + \bar{u}$, a belief that $\bar{u} > 0$ ($\bar{u} < 0$) implies that, for the given β , the observed mean return, \bar{r} , is higher (lower) than would have been expected, given one’s prior on α and the sample mean \bar{x} . In this way, we can think of assertions about \bar{u} as implicit statements about one’s conditional prior for α , given β .

This turns out to be a particularly convenient perspective, given our simulation methodology and the hypotheses of interest. For example, individuals with strong priors that expected return variation is an equilibrium response to changes in risk would likely have strong priors that expected real stock returns are always positive.⁶ On the other hand, one often sees statements in the financial press, and sometimes in the academic literature, that the market is substantially ‘overvalued’. Such individuals would presumably have strong priors that expected returns are sometimes negative. Given the apparent diversity of prior beliefs about this issue, it is interesting to investigate the conclusions that would be drawn from the given empirical evidence for a range of beliefs.

Specifying α as we have simplifies the analysis in that it allows us to formulate the hypothesis of nonnegative expected returns in terms of an upper bound on β . To see this, substitute $\bar{r} - \beta\bar{x}$ for α in the expected return expression, $\alpha + \beta x$, to get $\bar{r} + \beta(x - \bar{x})$. Assuming $\beta > 0$, the minimum expected return in a given sample period is $\bar{r} + \beta(x_{\min} - \bar{x})$. Setting this expression equal to zero and solving for β yields $\beta_0 = \bar{r}/(\bar{x} - x_{\min})$. Therefore, the hypothesis that expected returns are nonnegative during the given period is equivalent to $\beta \leq \beta_0$. Furthermore, the expected return corresponding to a slope of $2\beta_0$ is $-\bar{r}$. More generally, if $\bar{u} \neq 0$, then $(\bar{r} - \bar{u}) - \beta\bar{x} = \alpha$; hence, the analysis is modified by substituting $\bar{r} - \bar{u}$ for \bar{r} in the expressions for β_0 . Except where noted otherwise, we assume that $\bar{u} = 0$ in the analysis below.

A limitation of this approach is that it treats β_0 as if it were an exogenous constant. A fuller treatment of the endogeneity of expected returns in a dynamic context is left to future research.

7. Tests for expected real return variation

This section presents bootstrap evidence for the regressions of real equal-weighted and value-weighted index returns on B/M and dividend yield for the entire period, 1926–91, and the 1941–91 and 1963–91 subperiods.

⁶ It is possible to devise models in which the expected real market *premium* is negative (Boudoukh et al., 1996). Since real T-bill returns exceeded 4.5% in the deflationary period of the late 1920s and early 1930s, however, the negative premium required to make total expected real return negative during that period would have to be quite large.

Table 5

Regression of one-year CRSP equal-weighted index real returns on book-to-market or dividend yield: 1926–1991

β %	Mean b (%)	p -value	LKLHD	Mean b (%)	p -value	LKLHD
Book-to-market results				Dividend yield results		
0	1.88	0.001	0.001	1.23	0.127	0.041
1	2.77	0.002	0.003	2.20	0.167	0.047
2	3.83	0.004	0.004	3.32	0.213	0.067
3	4.91	0.002	0.003	4.37	0.279	0.075
4	5.91	0.007	0.008	5.39	0.352	0.084
5	6.83	0.004	0.010	6.27	0.426	0.095
6	7.85	0.010	0.012	7.26	0.518	0.100
7	8.79	0.014	0.017	8.28	0.617	0.101
8	9.77	0.022	0.031	9.31	0.692	0.092
9	10.81	0.026	0.036	10.24	0.771	0.080
10	11.80	0.040	0.046	11.10	0.836	0.069
11	12.79	0.062	0.064	12.43	0.902	0.051
12	13.84	0.075	0.089	13.29	0.937	0.036
13	14.64	0.103	0.104	14.24	0.955	0.025
14	15.81	0.149	0.148	15.23	0.974	0.016
15	16.80	0.202	0.198	16.49	0.988	0.009
16	17.77	0.257	0.227	17.26	0.990	0.005
17	—	—	—	18.25	0.996	0.003
18	—	—	—	19.30	0.998	0.002
19	—	—	—	20.36	0.999	0.001

One-year returns are regressed on beginning-of-year B/M or dividend yield (scaled by standard deviation) to obtain the historical slope estimate, and a first-order autoregression is estimated for B/M. Bias-adjusted estimates of the regression parameters are used to create adjusted regression residuals. Given a hypothetical value of β , the return regression slope, and the bias-adjusted autocorrelation estimate, artificial time series of annual returns and B/M or dividend yield ratios ($T = 66$) are generated by randomly sampling from the joint distribution (calendar year pairs) of adjusted residuals.

The (artificial) returns are regressed on the (artificial) B/M or dividend yield ratios to obtain an estimate of β . The procedure is repeated 2500 times and b , the average of the regression coefficient estimates, is computed.

The bootstrap p -value for a given β is the proportion of the 2500 estimates that exceed the historical slope. LKLHD (the bootstrap likelihood) is the scaled proportion of the 2500 estimates that lie within a small interval centered at the historical slope. The proportions are scaled so that the sum of likelihoods over all β 's is one.

7.1. Equal-weighted returns regressed on B/M or dividend yield

The bootstrap evidence from the regressions of equal-weighted index annual returns on B/M is reported in columns 2–4 of Table 5 for values of β from zero to 16%. The maximum is, apart from rounding, twice the value ($\beta = 8\%$) for which

the minimum expected return is zero. This ensures that the lowest implied expected return over 1926–91 is the negative of the sample mean return or -15.3% , as discussed above.

The bootstrap standard error is 4.3% under the null hypothesis that $\beta = 0$, and it varies very little as the true value of β is altered in the simulations. It is higher than, but close to, the OLS standard error, 4.1%. The OLS estimate is biased upward, however, as discussed earlier. The average simulated value of the OLS estimator given in the second column tends to exceed the true value by 1.8–1.9%, a bit higher than the 1.75% predicted using Stambaugh's formula.

The bootstrap p -value in the third column of Table 5, which is actually a simulation estimate of the p -value, is 0.001. This is the proportion of the 2500 simulated (under the null) estimates that exceed the historical slope of 20.2%. The standard error of this estimate, based on a binomial analysis, is the square root of $(0.001)(0.999)/2500$, which is less than 0.001. The standard error for a p -value of 0.01 ($\beta = 6$) is 0.002. Having strongly rejected the null hypothesis of a zero slope, we would like additional information regarding the true value of β . Examining the p -values in column 3, we see that values of β up to 6% can be rejected at the 0.01 level, while values up to 10% can be rejected at the 0.05 level. At this slope value, a move of one-standard deviation in B/M implies an increase in expected return of a whopping 1000 basis points.

We now shift the focus from asking which values of β can be rejected at conventional significance levels to which values have highest likelihood, conditional on the one-year historical slope estimate. The likelihoods reported in column 4 of Table 5 start at 0.001 for a slope of zero and increase (almost) monotonically to 0.23 at $\beta = 16\%$, the maximum considered. These likelihoods (or posterior probabilities under the uniform prior) are computed using a neighborhood consisting of points within 1% of the historical slope, and are plotted in Fig. 1 (see triangles). The values of beta with high likelihood correspond to huge levels of expected return variation. Under the (naive) uniform prior, most of the posterior probability exceeds $\beta = 8\%$, which is the largest value of beta that ensures that conditional expected returns (fitted values) are always non-negative over the 1926–91 period.

In general, the ratio of posterior probability for a null hypothesis to posterior probability for an alternative (the 'posterior odds ratio') equals the prior odds ratio, $p/(1-p)$, times another ratio often referred to as the 'Bayes factor' (see Jeffreys, 1957). This factor is the weighted average likelihood under the null hypothesis divided by the weighted average likelihood under the alternative, with weights determined by the prior. In a test of the null hypothesis that there is no variation in expected returns related to B/M, the Bayes factor is just the likelihood at $\beta = 0$ divided by a weighted average of the likelihoods for positive betas. From Fig. 1 it is clear that the Bayes factor will always be less than one (over the range of slopes considered) and it will be close to zero for any prior that puts substantial weight on the larger values of beta. Thus, we have a sharp shift

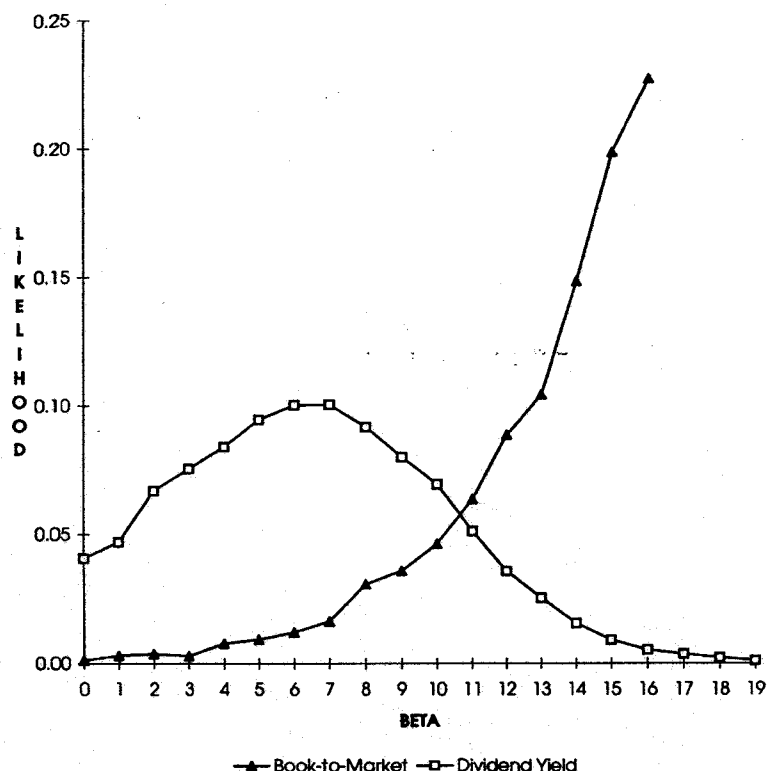


Fig. 1. Bootstrap likelihoods for equal-weighted index returns: 1926–1991

Beta is the slope coefficient in the regression of annual real returns on prior dividend yield or Dow book-to-market ratio, scaled so that beta gives the change in expected return for a one standard deviation change in the independent variable. The (unscaled) likelihood for a given beta is the proportion of 2500 bootstrap simulation estimates, generated under that specification of beta, which fall in a small interval around the historical slope estimate for the 1926–91 period. Likelihoods are scaled to sum to one.

in beliefs away from the hypothesis $\beta = 0$, particularly for those individuals who believe that expected returns are sometimes negative.

The bootstrap evidence from the regression of equal-weighted index returns on dividend yield is reported in columns 5–7 of Table 5. The maximum β values considered correspond to a minimum expected return of -15.3% in each case. The range of slopes for B/M is smaller than that for dividend yield since the minimum value of B/M is more extreme when measured in units of standard deviations from the mean. The bootstrap standard error is 4.9% under the null hypothesis $\beta = 0$. It is higher than, but again close to, the OLS standard error of 4.8% in Table 2. The OLS bias (difference between the first two columns) ranges

from 1.2% to 1.4% , comparable to the 1.3% predicted using Stambaugh's bias formula. The bias is roughly 20% of the value of the historical slope, 6.7 .

The bootstrap p -value for $\beta = 0$ is 0.13 , close to the p -value of 0.12 based on the t -statistic of 1.39 in Table 2. Although we cannot reject the null hypothesis of a zero slope at conventional levels, the likelihood function in Fig. 1 (see squares) rises initially, peaks at $\beta = 7\%$, and declines steadily after that. Thus, the Bayes factor will be less than one, i.e., will favor the conclusion that expected returns do vary, except for priors that place most of the weight on very large values of β (greater than 11%). Therefore, anyone who believes that expected returns were positive over the period 1926–91 ($\beta \leq 9\%$) will become more confident that expected returns vary with dividend yield. The maximum possible shift in odds is by a factor of $2.5 = 0.101/0.041$, the ratio of the maximum likelihood to the likelihood at zero. On the other hand, an investor with prior beliefs heavily weighted on β values greater than 11% will become less confident that expected returns vary with yield. This is a relative statement, though; such an investor will still be confident (high posterior probability) that $\beta \neq 0$, but this will be due to the prior dominating the sample evidence.

7.2. Value-weighted returns regressed on B/M or dividend yield

Bootstrap evidence for the value-weighted index is given in Table 6. The bootstrap p -values for $\beta = 0$ are close to 0.02 for both B/M and dividend yield, rejecting the null of no expected return variation. The likelihood functions plotted in Fig. 2 tell a subtler story, however. As earlier, the B/M likelihood function never drops below the $\beta = 0$ likelihood over the range of slopes considered, although it does turn downward. Thus, the evidence implies a shift in beliefs away from the hypothesis of no expected return variation for the entire class of priors considered. In contrast, while the likelihood function for dividend yield generally favors positive values of β , an investor with a prior concentrated on very large values of β would come away from the evidence less confident that expected return varies with dividend yield. This follows from the fact that the likelihood function for dividend yield eventually drops below the likelihood for $\beta = 0$, as earlier in Fig. 1.

7.3. Subperiod analyses

This section discusses bootstrap simulation results for the 1941–91 and 1963–91 subperiods. The 1941–91 subperiod excludes the Great Depression period of high return volatility; the more recent post-1962 subperiod is of interest because it coincides with the period over which much cross-sectional examination of book-to-market ratios has been undertaken.

Not surprisingly, given the smaller time-series sample and greater persistence in the independent variables over the subperiods, the estimated slope biases

Table 6
Regression of one-year CRSP value-weighted index real returns on book-to-market or dividend yield: 1926–1991

β %	Mean b (%)	p -value	LKLHD	Mean b (%)	p -value	LKLHD
Book-to-market results				Dividend yield results		
0	1.31	0.016	0.013	0.94	0.022	0.020
1	2.43	0.032	0.023	1.84	0.043	0.029
2	3.36	0.050	0.035	2.86	0.076	0.047
3	4.41	0.085	0.050	3.90	0.140	0.067
4	5.35	0.134	0.073	4.89	0.212	0.090
5	6.36	0.203	0.103	5.90	0.313	0.115
6	7.40	0.306	0.129	6.72	0.417	0.131
7	8.35	0.429	0.156	7.81	0.571	0.139
8	9.39	0.577	0.155	8.91	0.720	0.122
9	10.35	0.731	0.146	9.90	0.836	0.096
10	11.36	0.846	0.118	10.82	0.912	0.068
11	—	—	—	11.92	0.965	0.041
12	—	—	—	12.79	0.987	0.023
13	—	—	—	13.85	0.996	0.009
14	—	—	—	14.90	0.998	0.003
15	—	—	—	15.94	1.000	0.001
16	—	—	—	16.86	1.000	0.000

One-year returns are regressed on beginning-of-year B/M or dividend yield (scaled by standard deviation) to obtain the historical slope estimate, and a first-order autoregression is estimated for B/M. Bias-adjusted estimates of the regression parameters are used to create adjusted regression residuals. Given a hypothetical value of β , the return regression slope, and the bias-adjusted autocorrelation estimate, artificial time series of annual returns and B/M or dividend yield ratios ($T = 66$) are generated by randomly sampling from the joint distribution (calendar year pairs) of adjusted residuals.

The (artificial) returns are regressed on the (artificial) B/M or dividend yield ratios to obtain an estimate of β . The procedure is repeated 2500 times and b , the average of the regression coefficient estimates, is computed.

The bootstrap p -value for a given β is the proportion of the 2500 estimates that exceed the historical slope. LKLHD (the bootstrap likelihood) is the scaled proportion of the 2500 estimates that lie within a small interval centered at the historical slope. The proportions are scaled so that the sum of likelihoods over all β 's is one.

from Eq. (3) are larger for the subperiod estimators. For example, the simulated bias for the B/M slope (not reported in tables) exceeds 4.0 for both indexes over 1963–91. Bias-adjusted t -statistics are computed by subtracting the simulated mean value (under the null $\beta = 0$) from the estimator b and dividing by the standard deviation of the bootstrap slopes.

Table 7 presents bootstrap p -values for $\beta = 0$, along with p -values based on the OLS t -statistics and bias-adjusted t -statistics. Results for 1941–91 are given

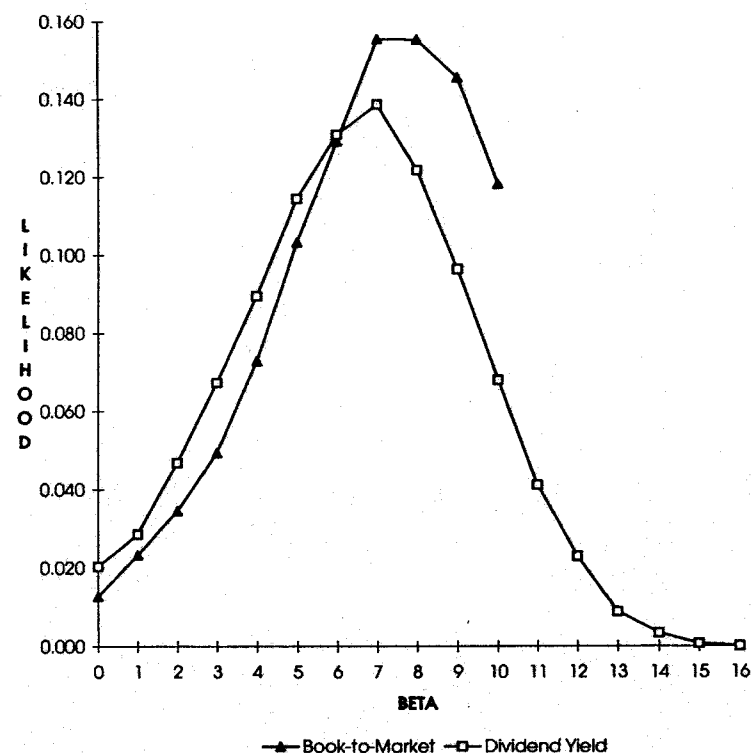


Fig. 2. Bootstrap likelihoods for value-weighted index returns: 1926–1991

Beta is the slope coefficient in the regression of annual real returns on prior dividend yield or Dow book-to-market ratio, scaled so that beta gives the change in expected return for a one standard deviation change in the independent variable. The (unscaled) likelihood for a given beta is the proportion of 2500 bootstrap simulation estimates, generated under that specification of beta, which fall in a small interval around the historical slope estimate for the 1926–91 period. Likelihoods are scaled to sum to one.

in panel A, while results for 1963–91 are in panel B. As expected, the bias-adjusted p -values are often much larger than the conventional OLS p -values. However, when the values are relatively low, the bias-adjusted p -values are smaller than the bootstrap p -values. This apparently reflects nonnormality of the bootstrap distribution for the slope estimator, with the bootstrap having a fatter upper tail. For example, the adjusted p -value is 0.008 in the equal-weighted dividend yield case (1941–91), while the bootstrap p -value is 0.048. The binomial standard errors are 0.002 for the 0.008 p -value and 0.004 for the 0.048 p -value.

We focus on the more conservative bootstrap p -values. In the post-1940 period, the hypothesis of no related expected return variation is rejected at the

Table 7
P-values for the null hypothesis $\beta = 0$: subperiods 1941–91 and 1963–91

Index	Financial ratio	OLS	Bias-adjusted	Bootstrap
Panel A: 1941–91				
Equal-weighted	Book-to-market	0.002	0.014	0.040
Equal-weighted	Dividend yield	0.001	0.008	0.048
Value-weighted	Book-to-market	0.031	0.127	0.156
Value-weighted	Dividend yield	0.001	0.008	0.024
Panel B: 1963–91				
Equal-weighted	Book-to-market	0.036	0.295	0.250
Equal-weighted	Dividend yield	0.001	0.012	0.023
Value-weighted	Book-to-market	0.156	0.629	0.571
Value-weighted	Dividend yield	0.008	0.102	0.115

Equal-weighted index is the CRSP equal-weighted annual real return inclusive of dividends from April of year t to March of year $t + 1$.

Value-weighted index is the CRSP value-weighted annual real return inclusive of dividends from April of year t to March of year $t + 1$.

Book-to-market is the ratio of the book value per share at the end of year $t - 1$ of the stocks in the Dow Jones Industrial Average to the Dow Jones Industrial Average at the end of March of year t .

Equal-weighted dividend yield is the ratio of dividends paid from April of year $t - 1$ to March of year t to the accumulated value of one dollar invested in the CRSP equal-weighted index exclusive of dividends for one year ending on March 31 of year t .

Value-weighted dividend yield is the ratio of dividends paid from April of year $t - 1$ to March of year t to the accumulated value of one dollar invested in the CRSP value-weighted index exclusive of dividends for 1-year ending on March 31 of year t .

0.05 level, except for the combination of value-weighted index and B/M. The smallest p -value is 0.024 for the regression of value-weighted returns on dividend yield. In the post-1962 period, rejection at the 0.05 level is only observed for the equal-weighted index and dividend yield.

The bootstrap likelihood functions provide additional insights into these results, as well as the potential contrasts between classical and Bayesian inference. In all but one case, an investor who firmly believes that expected returns are never negative will come away from the regression evidence with greater confidence that expected returns vary with B/M or dividend yield. This is true even for the post-1962 scenario with B/M and the equal-weighted index (with bootstrap p -value 0.25), although the maximum shift in odds (Bayes factor) is only 1.7 in this case. On the other hand, in most cases, the direction of the shift in beliefs for an investor who firmly believes that expected returns are sometimes negative depends on the specifics of their prior distribution.

The post-1962 likelihood function for the equal-weighted index and dividend yield rises steeply and stays well above the likelihood at $\beta = 0$ for the entire range of β values. This will be viewed by *all* investors (for the class of priors considered) as evidence in support of time-varying expected returns. At the other end of the spectrum, the value-weighted, B/M, post-1962 likelihood function for the value-weighted index and B/M steadily declines as β increases, shifting all priors toward the conclusion that B/M does not track variation in expected returns for the value-weighted index. One final case worth highlighting is the post-1962 scenario for the value-weighted index and dividend yield. Here, the likelihood function drops below the likelihood at $\beta = 0$ only at the largest value considered ($\beta = 9\%$). Thus, virtually all investors considered will view the evidence as supporting a relation between dividend yield and expected return, even though the classical test fails to reject the null hypothesis of no relation at the 0.10 level.

8. Tests for nonnegative expected returns

The regression evidence examined above suggests considerable variation in expected returns as a function of both B/M and dividend yield. Fig. 3 plots the

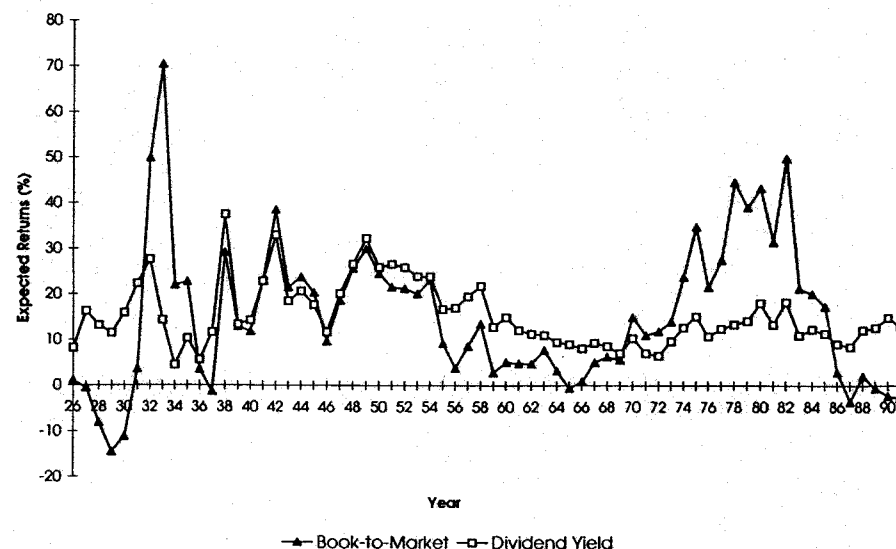


Fig. 3. Expected equal-weighted index returns: 1926–1991

Expected returns are fitted values from the regression of real returns on prior book-to-market ratio or dividend yield over the 1926–1991 period. The fitted values are evaluated at the maximum likelihood beta values, 16% for book-to-market and 6% for dividend yield, with the intercepts specified so as to equate the average expected returns to the sample mean return.

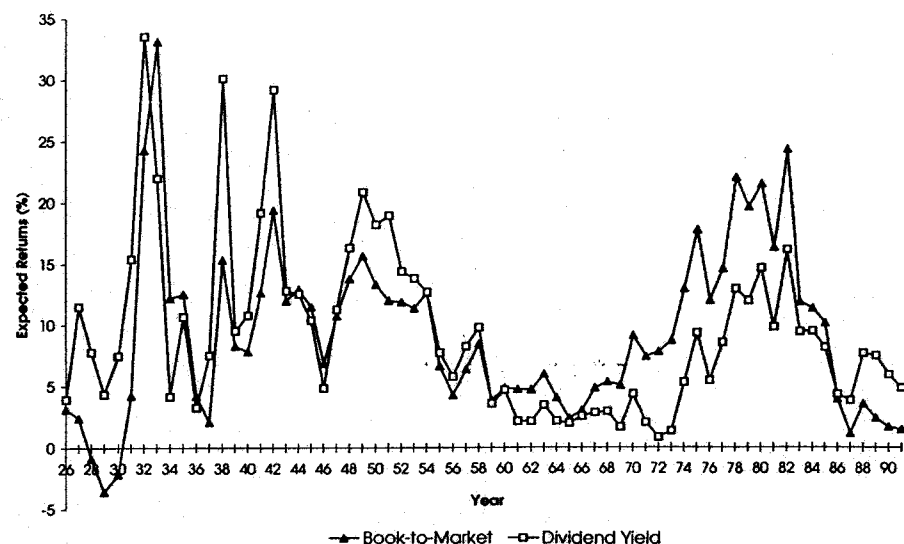


Fig. 4. Expected value-weighted index returns: 1926–1991

Expected returns are fitted values from the regression of real returns on prior book-to-market ratio or dividend yield over the 1926–1991 period. The fitted values are evaluated at the maximum likelihood beta values, 16% for both book-to-market and dividend yield, with the intercepts specified so as to equate the average expected returns to the sample mean return.

1926–91 equal-weighted expected returns (fitted values) using historical B/M or dividend yield and the maximum likelihood slope values $\beta = 16\%$ and $\beta = 7\%$, respectively. Consistent with the market being predictably overvalued, the expected returns based on B/M are negative for several years at both the beginning and end of the sample and two years in between. The most negative expected return is -14.5% in 1929 when the market value for the Dow 30 is about 3.7 (1/0.27) times book value. Expected returns based on dividend yield are always positive, however. Value-weighted index expected returns are plotted in Fig. 4. Here, negative expected returns based on B/M are smaller in magnitude and limited to the 1928–30 period.

Perhaps as implausible as negative expected returns, at least to a believer of market efficiency, are the large positive expected returns implied by the high B/M values, particularly from the mid-1970s to the early 1980s. B/M values in this period imply expected real returns ranging from 20% to 50% per year. While these rates are extremely high, an interpretation that the market was on the brink of a near-term upward correction is too simplistic. Anyone attempting to capitalize on the 'undervaluation' in the 1970s would have had to endure a prolonged wait before the market began to rise in the early 1980s. A similar statement can be made regarding the market's performance since the late 1980s.

The Dow Jones B/M has been about 0.4 during this period, fairly low, by historical standards. However, expectations of a low market return have been foiled by the stellar market performance in the 1990s to date. The Dow Jones B/M stood at its historical lowest, 0.27, on April 1, 1995, but the market performance since then has been one of the best in the post-World War II period. The Dow Jones B/M is currently (August 1996) only 0.23.

In the remainder of this section, we consider formal tests of the hypothesis that expected real returns are always non-negative over the given period of interest. By the law of iterated expectations, if expected returns are non-negative conditional on all information available at each point in time, then expected returns must also be non-negative conditional on any subset, in particular when conditioning on B/M or dividend yield.

8.1. Classical tests

From Table 5, we see that the hypothesis that expected returns are non-negative over the 1926–91 period is rejected at the 0.02 level (p -value for $\beta \leq 8$) for B/M and equal-weighted index returns. The corresponding p -value for B/M and value-weighted index returns in Table 6 is about 0.20 (for $\beta \leq 5$). Thus, we cannot reject, at conventional significance levels, the hypothesis that the negative fitted values for the value-weighted index, in Fig. 4, are due to error in estimating β . There is no evidence of negative expected returns in the dividend yield results.

As explained in Section 6, these non-negativity tests implicitly assume that the sample mean of the return disturbances, \bar{u} , is zero for the given period. We now compare this test to the conventional regression procedure for making inferences about a conditional mean. The estimated expected return (fitted value) at the minimum sample value of B/M (27.2% for 1929) is -22.3% for the equal-weighted index. The corresponding standard error, based on the usual regression formula, is 17.3%. This yields a t -statistic of -1.3 and a p -value of 0.10 for the hypothesis that the minimum expected return is positive over the 1926–91 period.

The fairly large increase in p -value (0.02 to 0.10) reflects the fact that the conventional procedure incorporates uncertainty about the sample mean disturbance term, whereas our β -based test conditions on a belief about the magnitude of \bar{u} .⁷ Due to the substantial variability in return surprises, the conventional test for non-negative expected returns is not powerful. Conditioning on the mean disturbance permits us to concentrate on the slope coefficient and the implications of time variation for nonnegative expected returns, holding the average level of expected return constant at the sample mean. Since we focus on a narrower issue, the power of our test is, not surprisingly, enhanced.

⁷ A more formal statistical appendix is available from the authors on request.

Now, suppose that (relative to one's priors) the data indicate that the mean return surprise, \bar{u} , is positive. For example, the literature on the equity premium puzzle (Mehra and Prescott, 1985) suggests a lower prior expected real return on the equal-weighted index than the sample mean of 15.3%. Alternatively, one might believe that the observed sample of U.S. returns is subject to a selection bias related to survival and growth of equity markets at the country level (Brown et al., 1995).

Suppose, for example, that half of the mean return is viewed as an ex post surprise, i.e., $\bar{u} = 7.65\%$. Given the analysis in Section 6, this has the effect of cutting in half the maximum value of β consistent with non-negative expected returns. The associated p -value (for $\beta \leq 4$) is 0.002, as compared to 0.02 (for $\beta \leq 8$) earlier. Naturally, incorporating a belief that the return surprises are positive, on average, leads to a stronger rejection of non-negativity, since our belief about the general level of expected returns is lowered relative to that of the previous scenario.

8.2. Bayesian tests

We now turn to Bayesian tests of the nonnegativity restriction. The equal-weighted B/M regression is examined first. In order to analyze the expected return hypothesis in a manageable way that is readily communicated to a variety of readers, some additional structure is needed. We consider priors that assign probability p to betas from zero to 8% and $1 - p$ to betas from 9% to 16%, the prior probability distributed uniformly within each subset. With this prior, p is the probability for the composite null hypothesis that expected returns are always non-negative; conditional on this hypothesis, no variation in expected returns ($\beta = 0\%$) is considered as plausible as anything up to an 800-basis-point ($\beta = 8\%$) variation in expected returns for every one-standard-deviation change in B/M. Prior weight on beta values between 9% and 16% entails a belief that expected returns can be negative (i.e., when B/M is relatively low) and that expected returns vary greatly through time.

The assumption that probability within each subset is uniformly distributed is adopted for simplicity. While it violates desirable conditions like continuity of the prior density, we believe the general conclusions that emerge are sufficiently robust that the benefits of simplification outweigh the costs. Given the uniformity assumption, the Bayes factor is just the ratio of average likelihood under the null to average likelihood under the alternative, which is independent of p . The Bayes factor for our equal-weighted index B/M regression is 0.086. Thus, in light of the B/M evidence, there is roughly a 12-to-1 shift in odds against the hypothesis that expected returns are always non-negative over the period 1926–91.

In Fig. 5, we plot the posterior probability that expected returns are always nonnegative as a function of the prior probability p . The 'no-change' line with a slope of one is a natural benchmark for which the prior and posterior probabilities are equal. Thus, the straight line in Fig. 5 corresponds to a scenario

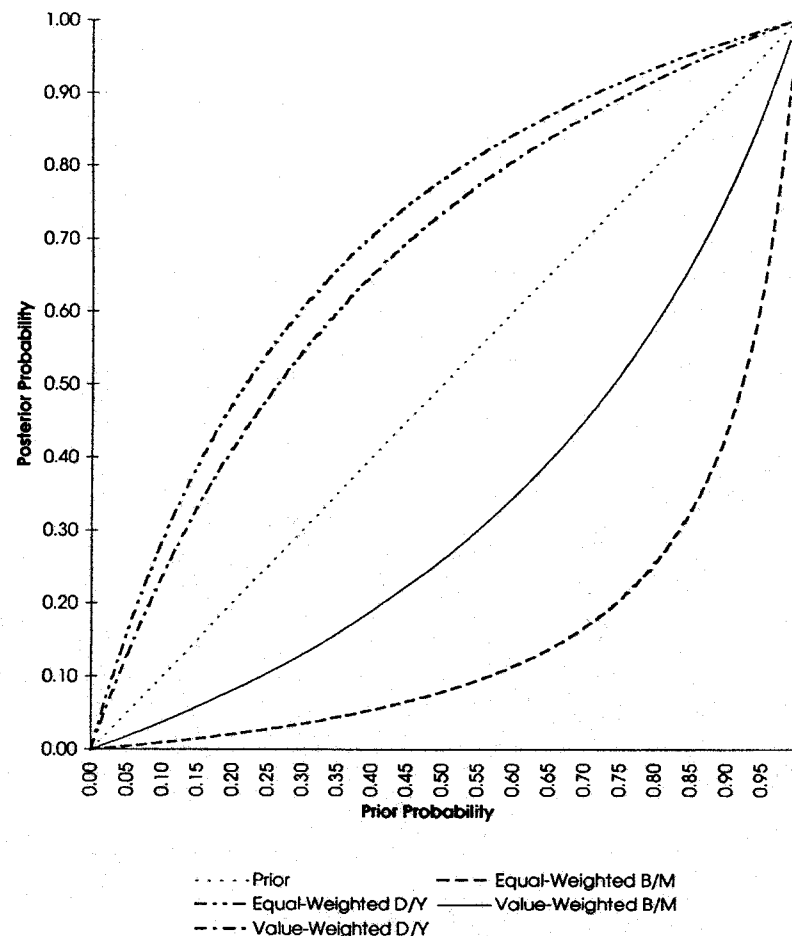


Fig. 5. Posterior probability that expected real returns are non-negative: 1926–1991
Posterior probabilities are derived from the likelihood functions plotted in Figs. 1 and 2, with prior distributions specified over the ranges of beta values in those figures. Beta values in the first half of each range are consistent with non-negative expected returns, while betas in the second half yield negative expected returns at some point in the 1926–1991 period. The prior distributions are uniformly distributed within each half; prior probability is the weight assigned to the first half of the beta range.

in which the Bayes factor is one. The low value of the actual Bayes factor in the equal-weighted B/M case ensures that the posterior probability curve lies below the no-change line for p between zero and one. This means that, regardless of the prior ($p \neq 0, 1$), one ends up less confident that expected returns are always non-negative (i.e., $0 \leq \beta \leq 8\%$). In particular, an individual who is initially

uncommitted ($p = 0.5$) with regard to the hypothesis that expected returns are non-negative has posterior probability 0.08, while an individual with prior confidence in the non-negativity condition, say $p = 0.8$, has posterior probability 0.26.

Now, as we did for the classical test of non-negativity in the previous subsection, we consider the alternative assumption that \bar{u} , the mean surprise component of return, is 7.65%. As before, the maximum β consistent with non-negative expected returns is reduced to 4%. In addition, the range of β 's entertained as plausible a priori is likewise cut in half, with a maximum β of 8%. At this value, the most negative expected return over the 1926–91 period is now $-(\bar{r} - \bar{u}) = -7.65\%$. Given the likelihoods reported in Table 5, the resulting Bayes factor (ratio of average likelihoods from 0 to 4 and 5 to 8) is 0.226, almost three times the Bayes factor obtained under the $\bar{u} = 0$ assumption. The corresponding posterior probability for non-negative expected returns of the 'uncommitted' ($p = 0.5$) investor is now 0.18, as compared to 0.08 above. Truncation has a minimal effect on this result.

The increase in posterior probability may be surprising, since the classical test rejected non-negativity more strongly when $\bar{u} = 7.65\%$. The reason is that the prior associated with $\bar{u} = 7.65\%$ eliminates the weight previously placed on the very large values of β (> 8) that have extremely large likelihood, given the historical slope of 20.2% (see Fig. 1). Thus, the increase is a consequence of the change in prior implied by the given belief about \bar{u} (maximum β reduced from 16% to 8%), in conjunction with the particular behavior of the likelihood function in this example.

In the value-weighted B/M case, the posterior probability curve again lies below the no-change line, despite the relatively large p -value of 0.20 (for $\beta \leq 5\%$ in Table 6) for the null hypothesis of non-negative expected returns. The posterior probabilities are less extreme than those for the equal-weighted index, however. For example, an observer with $p = 0.5$ comes away from the B/M evidence with posterior probability 0.25 that expected returns are non-negative. Posterior probability curves for dividend yield in Fig. 5 lie above the no-change line, reflecting increased confidence in the non-negativity hypothesis. Curves for the 1941–91 subperiod, given in Fig. 6, all lie above the no-change line as well.

8.3. Tests with truncation

One view of the 1933 outlier is that it is an observation that just happened to occur in one year out of 66 in our sample, but which has an ex ante (population) probability far less than this. From this perspective, conventional weighted least squares procedures do not adequately address the problem since they still give too much weight to the 1933 data point. To gain additional insight into the problem, we rerun the 1926–91 B/M bootstrap analysis with 1933 truncated, as discussed earlier in Section 3.

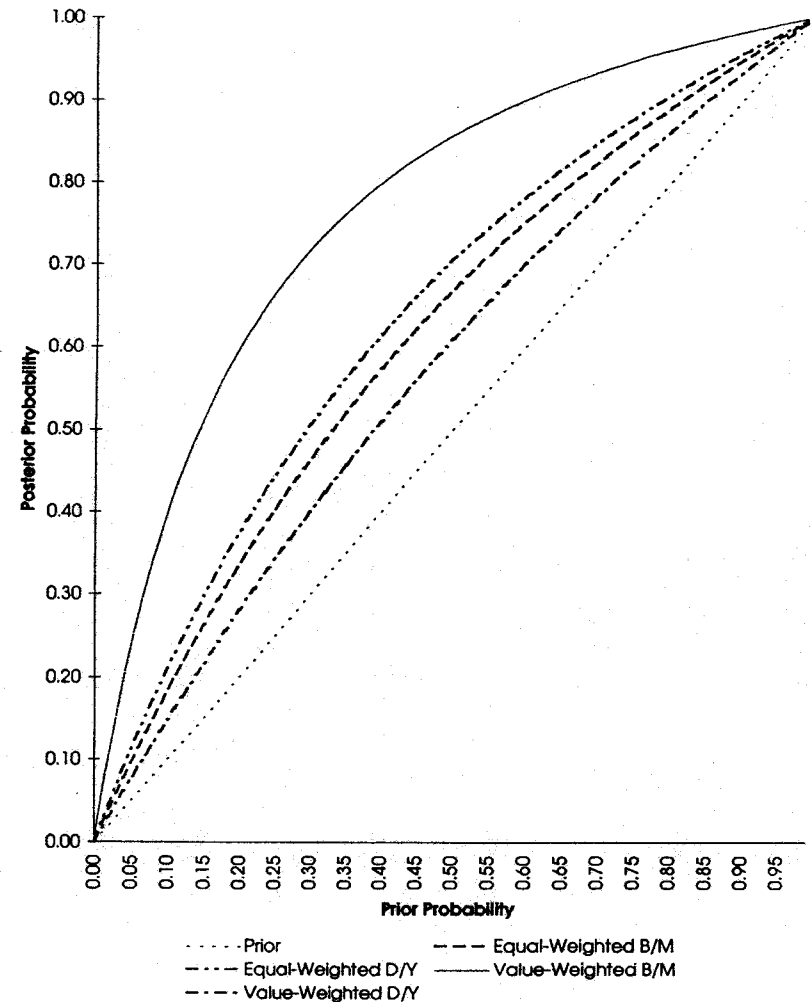


Fig. 6. Posterior probability that expected real returns are non-negative: 1941–1991. Posterior probabilities are derived from the likelihood functions and prior distributions specified over the corresponding ranges of beta values. Beta values in the first half of each range are consistent with non-negative expected returns, while betas in the second half yield negative expected returns at some point in the 1941–1991 period. The prior distributions are uniformly distributed within each half; prior probability is the weight assigned to the first half of the beta range.

With the equal-weighted stock index, the bootstrap p -value for the null hypothesis that expected returns are unrelated to B/M ($\beta = 0$) rises to 0.005 with truncation, from 0.001 earlier; the p -value for the hypothesis of non-negative expected returns ($\beta \leq 8$) increases to 0.084, from 0.022 when 1933 is included

(Table 5). Corresponding p -values for the value-weighted index are 0.045 (0.016 earlier) for no relation, and 0.27 (0.20 earlier) for non-negative expected returns.

Perhaps more informative, given the limitations of p -values, are the posterior probabilities. Now, an uncommitted, i.e., $p = 0.5$, observer comes away from the equal-weighted B/M evidence with posterior probability 0.14 (0.08 earlier) that expected returns were non-negative over the 1926–91 period. For the value-weighted index, the posterior probability is 0.34 (0.25 earlier). Thus there is still a shift in beliefs, albeit a weaker one, away from the non-negative expected returns hypothesis.

9. Summary and conclusions

We find reliable evidence that both dividend yield and B/M track time-series variation in expected real one-year stock returns over the period 1926–91 and the subperiod 1941–91. The B/M relation is stronger over the full period, while the dividend yield relation is stronger in the subperiod.

The estimated 1926–91 equal-weighted return slope on B/M implies that expected return moves by an astounding 20 percentage points (18 bias-adjusted) for a one-standard-deviation change in B/M. The coefficient is so large that implied expected returns are sometimes negative, particularly in the late 1920s when the B/M ratio is relatively low. The results are quite sensitive to the 1933 data point, however, when both B/M and market index returns attain their maximum values. Using a ‘truncation’ method that reduces the impact of this outlier, the p -value for non-negative expected returns is 0.08 (0.02 without truncation). Although the p -value of 0.08 might be interpreted by some as ‘insignificant’, the posterior probability for an investor who assigns prior probability 0.5 to the hypothesis of non-negative expected returns is only 0.14 (0.08 without truncation). Results for the value-weighted index and B/M are qualitatively similar, though the shifts in beliefs are far less dramatic.

Thus, the B/M results suggest that expected return variation over the 1926–91 period was not driven entirely by equilibrium changes in compensation for risk. Rather, it appears that the market may have been inefficient, particularly in the late 1920s and early 1930s, a period of great economic volatility. The stronger results for the equal-weighted index, which is influenced more by small-firm returns, is consistent with this conclusion, although greater variation in risk-related predictability is also plausible for smaller firms.

The market inefficiency conclusion should be tempered by several observations. First, there is the familiar element of data mining which implies that the significance of extreme results is overstated, although by how much it is difficult to say. While the financial ratios considered here can be motivated theoretically, dividend yield is examined in part because of its prominence as compared to, say, earnings yield in past time-series studies (e.g. Fama and French, 1988). B/M

is considered primarily because of its recently acquired celebrity status in explaining cross-sectional stock return variation, although the extent to which this biases time-series results at the market level is less clear. These concerns and the fact that the forecasting power of B/M and dividend yield varies considerably over different subperiods give further credence to the idea that data mining considerations should temper one’s assessment of the value of these ratios in making future investment decisions.

The dividend yield slope estimates for the 1926–91 period and all of the estimates for the 1941–91 subperiod are more ‘reasonable’ than the B/M results, but still substantial, implying expected return movements from three to eight percentage points for one-standard-deviation changes in the financial ratios. These estimates all result in varying degrees of *increased* confidence in the proposition that expected returns were never negative. While the considerable estimated variation in expected returns might be difficult to reconcile with market efficiency, the hypothesis that the slope is greater than three is never rejected at the 0.05 level. Definitive statements concerning the extent to which estimated variation in expected returns is related to risk or systematic mispricing (or both) require further analysis.

One step toward reconciliation with market efficiency might be based on the argument of Brown et al. (1995), who emphasize the role of survivor biases, even at the country level. These arguments gain credibility in that the evidence of inefficiency is tied to a historic period of considerable uncertainty. However, while survival of the U.S. market could impart an upward bias to the regression slope coefficient, it might also bias the intercept upward. Adjusting for the effect of survivor bias on the slope would reduce the statistical significance (larger p -value) of our evidence of negative expected returns, but the intercept bias would imply a stronger rejection of non-negativity, as shown in Section 8.1.

Violation of the homoskedastic linear regression specification used in most of our analysis might give a misleading impression as to the likelihood of negative expected returns, but our initial analysis suggests otherwise. The first-order autoregressive processes postulated for the financial ratios could easily be generalized, although we suspect this would have little impact on the main conclusions. It would also be interesting to compare the performance of the bootstrap likelihood methodology with the likelihood function generated under the assumption of bivariate normality of the vector autoregressive disturbances (see Kandel and Stambaugh, 1996). Although extension of the latter approach to overlapping multiyear returns would appear difficult, use of the bootstrap methodology in this context is straightforward and should provide further insights into the behavior of expected return variation.

In addition to our empirical conclusions, we note several examples of the limited informativeness of standard p -value analysis and the benefits of exploring the likelihood function. One interesting scenario illustrates the important

role that prior beliefs can sometimes play in interpreting data. An investor who is confident that expected returns are always positive views the regression results as supporting the hypothesis of expected return variation, while an investor strongly inclined to believe that expected returns are sometimes negative views the regression as favoring the constant expected return conclusion. In another scenario, investors spanning a wide variety of prior beliefs view the regression evidence as supporting a relation between dividend yield and expected return, even though we fail to reject the null hypothesis of no relation at the 0.10 level (p -value 0.12).

Finally, we note that the bootstrap approach to likelihood/Bayesian analysis is quite flexible and could prove useful in a variety of empirical contexts. This is likely to be the case in accounting and corporate finance applications, where the variables of interest are often far less well-behaved statistically than stock returns.⁸

References

- Berger, J., 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.
- Boudoukh, J., Richardson, M., Whitelaw, R., 1997. Nonlinearities in the relation between the equity risk premium and the term structure. *Management Science*, forthcoming.
- Breen, W.J., Korajczyk, R.A., 1994. On selection biases in book-to-market based tests of asset pricing models. Working paper no. 167, Northwestern University, Evanston, IL.
- Brown, S., Goetzmann, W., Ross, S., 1995. Survival. *Journal of Finance* 50, 853–873.
- Campbell, J., Shiller, R., 1988. The dividend price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.
- Chan, L.K.C., Jegadeesh, N., Lakonishok, J., 1995. Evaluating the performance of value versus glamour stocks: the impact of selection bias. *Journal of Financial Economics* 38, 269–296.
- Davis, J.L., 1994. The cross-section of realized stock returns: the pre-COMPUSTAT evidence. *Journal of Finance* 49, 1579–1593.
- DeGroot, M., 1982. Comment on Lindley's paradox. *Journal of the American Statistical Association* 77, 336–339.
- Dixon, W., Massey, F., 1969. *Introduction to Statistical Analysis*. McGraw-Hill, New York.
- Fama, E.F., French, K.R., 1988a. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–25.
- Fama, E.F., French, K.R., 1988b. Permanent and transitory components of stock prices. *Journal of Political Economy* 96, 246–273.
- Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25, 23–49.
- Fama, E.F., French, K.R., 1992. The cross-section of expected returns. *Journal of Finance* 47, 427–465.
- Fama, E.F., French, K.R., 1993. Common risk factors in the returns on bonds and stocks. *Journal of Financial Economics* 33, 3–56.

⁸The bootstrap programs used in our computations were written in Gauss and are available on request.

- Fama, E.F., Schwert, G.W., 1977. Asset returns and inflation. *Journal of Financial Economics* 5, 115–146.
- Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. *Journal of Finance* 48, 663–679.
- Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357–386.
- Jeffreys, H., 1957. *Theory of Probability*, 3rd ed. Oxford University Press, London.
- Kandel, S., McCulloch, R., Stambaugh, R.F., 1995. Bayesian inference and portfolio efficiency. *Review of Financial Studies* 8, 1–53.
- Kandel, S., Stambaugh, R.F., 1996. On the predictability of stock returns: an asset-allocation perspective. *Journal of Finance* 51, 385–424.
- Keim, D.B., Stambaugh, R.F., 1986. Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17, 357–390.
- Kendall, M.G., 1954. Note on bias in estimation of auto-correlation. *Biometrika* 41, 403–404.
- Kothari, S.P., Shanken, J., 1992. Stock return variation and expected dividends: a time-series and cross-sectional analysis. *Journal of Financial Economics* 31, 177–210.
- Kothari, S.P., Shanken, J., Sloan, R.G., 1995. Another look at the cross-section of expected returns. *Journal of Finance* 50, 185–224.
- Lindley, D., 1957. A statistical paradox. *Biometrika* 44, 187–192.
- Litzenberger, R.H., Ramaswamy, K., 1979. The effect of personal taxes and dividends on capital asset prices: theory and empirical evidence. *Journal of Financial Economics* 7, 163–195.
- Loughran, T., 1996. Is there a book-to-market effect? Working paper, University of Iowa, Iowa City, IA.
- MacBeth, J., Emanuel, D., 1993. Tactical asset allocation: pros and cons. *Financial Analysts Journal* November–December, 30–43.
- Mankiw, N.G., Shapiro, M.D., 1986. Do we reject too often? Small sample properties of tests of rational expectations models. *Economic Letters* 20, 139–145.
- Markowitz, H., Usmen, N., 1996. The likelihood of various stock market return distributions, Part 1: principles of inference. *Journal of Risk and Uncertainty* 13, 207–219.
- Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. *Journal of Monetary Economics* 15, 145–161.
- Nelson, C.R., Kim, M.J., 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance* 48, 641–661.
- Officer, R.R., 1973. The variability of the market factor of the New York Stock Exchange. *Journal of Business* 46, 434–453.
- Pontiff, J., Schall, L., 1995. Book-to-market as a predictor of market returns. Working paper, University of Washington, Seattle.
- Rozeff, M., 1984. Dividend yields are equity risk premiums. *Journal of Portfolio Management* 11, 68–75.
- Schwert, G.W., 1989. Why does stock market volatility change over time? *Journal of Finance* 44, 1115–1153.
- Shanken, J., 1987a. Multivariate proxies and asset pricing relations: living with the roll critique. *Journal of Financial Economics* 18, 91–110.
- Shanken, J., 1987b. A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics* 19, 195–215.
- Shiller, R.J., 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity* 2, 457–498.
- Stambaugh, R.F., 1986. Bias in regressions with lagged stochastic regressors. Unpublished manuscript, University of Chicago, Chicago, IL.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. *Econometrica* 48, 817–838.