



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Financial Economics 74 (2004) 209–235

JOURNAL OF
Financial
ECONOMICS

www.elsevier.com/locate/econbase

Predicting returns with financial ratios[☆]

Jonathan Lewellen*

MIT Sloan School of Management, Cambridge, MA 02142, USA

Received 9 January 2001; accepted 20 November 2002

Available online 18 May 2004

Abstract

This article studies whether financial ratios like dividend yield can predict aggregate stock returns. Predictive regressions are subject to small-sample biases, but the correction used by prior studies can substantially understate forecasting power. I show that dividend yield predicts market returns during the period 1946–2000, as well as in various subsamples. Book-to-market and the earnings-price ratio predict returns during the shorter sample 1963–2000. The evidence remains strong despite the unusual price run-up in recent years.

© 2004 Elsevier B.V. All rights reserved.

JEL classification: C32; G12

Keywords: Predictive regressions; Bias; Expected returns; Equity premium

1. Introduction

Fifty years ago, [Kendall \(1953\)](#) observed that stock prices seem to wander randomly over time. Kendall, and much of the early literature on market efficiency, tested whether price changes could be predicted using past returns. Empirical tests later expanded to other predictive variables, including interest rates, default spreads, dividend yield, the book-to-market ratio, and the earnings-price ratio (e.g., [Fama](#)

[☆]I am grateful to John Campbell, Ken French, Will Goetzmann, Sydney Ludvigson, Bill Schwert, Jay Shanken, Rob Stambaugh, Tuomo Vuolteenaho, and especially two anonymous referees for helpful comments and suggestions. Financial support from the MIT/Merrill Lynch Partnership is gratefully acknowledged.

*Tel.: +1-617-258-8408; fax: +1-617-258-6855.

E-mail address: lewellen@mit.edu (J. Lewellen).

and Schwert, 1977; Campbell, 1987; Fama and French, 1988; Campbell and Shiller, 1988; Kothari and Shanken, 1997).

The three financial ratios— DY , B/M , and E/P —share several common features. First, the ratios all measure stock prices relative to fundamentals. Because each has price in the denominator, the ratios should be positively related to expected returns. According to the mispricing view, the ratios are low when stocks are overpriced; they predict low future returns as prices return to fundamentals. The rational-pricing story says, instead, that the ratios track time-variation in discount rates: the ratios are low when discount rates are low, and high when discount rates are high; they predict returns because they capture information about the risk premium. DY , B/M , and E/P also share similar time-series properties. At a monthly frequency, they have autocorrelations near one and most of their movement is caused by price changes in the denominator. These statistical properties have a big impact on tests of return predictability.

This article provides a new test of whether the financial ratios can predict aggregate stock returns. I focus primarily on DY because it has received the most attention in the literature, and I focus exclusively on short-horizon tests—monthly returns regressed on lagged DY —to avoid the complications arising from overlapping returns. Prior studies find weak evidence, at best, that DY can forecast market returns. Fama and French (1988) find that DY predicts monthly NYSE returns from 1941–1986, with t -statistics between 2.20 and 3.21 depending on the definition of returns (equal- vs. value-weighted; real vs. nominal). However, Stambaugh (1986) and Mankiw and Shapiro (1986) show that predictive regressions can be severely biased toward finding predictability. Nelson and Kim (1993) replicate the Fama and French tests, correcting for bias using bootstrap simulations, and estimate that the p -values are actually between 0.03 and 0.33. More recently, Stambaugh (1999) derives the exact small-sample distribution of the slope estimate assuming that DY follows a first-order autoregressive (AR1) process. He reports a one-sided p -value of 0.15 when NYSE returns are regressed on DY over the period 1952–1996.¹

In this article, I show that the small-sample distribution studied by Stambaugh (1986, 1999) and Nelson and Kim (1993)—which has become standard in the literature—can substantially understate, in some circumstances, DY 's predictive ability. Although the bias correction is generally appropriate, we can sometimes improve upon it using knowledge of DY 's autocorrelation. The slope in a predictive regression is strongly correlated with DY 's sample autocorrelation, so any information conveyed by the autocorrelation helps produce more powerful tests of predictability. Incorporating this information into empirical tests has two effects: (1) the slope estimate is often larger than the standard bias-adjusted estimate; and (2) the variance of the estimate is much lower. In combination, the two effects can substantially raise the power of empirical tests.

¹ DY predicts long-horizon returns more strongly, but the statistical significance is sensitive to the time period considered and small-sample corrections. See, for example, Hodrick (1992), Goetzmann and Jorion (1993), Nelson and Kim (1993), and Ang and Bekaert (2002).

To gain some intuition, consider the model of returns analyzed by Stambaugh (1986, 1999), Mankiw and Shapiro (1986), and Nelson and Kim (1993):

$$r_t = \alpha + \beta x_{t-1} + \varepsilon_t, \quad (1a)$$

$$x_t = \phi + \rho x_{t-1} + \mu_t, \quad (1b)$$

where r_t is the stock return and x_{t-1} is the dividend yield (or other financial ratio). Eq. (1a) is the predictive regression and Eq. (1b) specifies an AR1 process for DY . The residuals, ε_t and μ_t , are correlated because positive returns lead to a decrease in DY . As a consequence, estimation errors in the two equations are closely connected:

$$\hat{\beta} - \beta = \gamma(\hat{\rho} - \rho) + \eta, \quad (2)$$

where η is a random error with mean zero and γ is a negative constant (described further below). Empirical tests are typically based on the marginal distribution of $\hat{\beta}$ from Eq. (2), integrating over all possible values of $\hat{\rho} - \rho$ and η . For example, the bias in $\hat{\beta}$ is found by taking expectations of both sides; the well-known downward bias in $\hat{\rho}$ induces an upward bias in $\hat{\beta}$ (since γ is negative). Notice, however, that this approach implicitly discards any information we have about $\hat{\rho} - \rho$. In particular, if we are willing to assume that DY is stationary, then a lower bound on the sampling error in ρ is $\hat{\rho} - 1$. In turn, Eq. (2) implies that the ‘bias’ in $\hat{\beta}$ is at most $\gamma(\hat{\rho} - 1)$. This upper bound will be less than the standard bias-adjustment if $\hat{\rho}$ is close to one, and empirical tests that ignore the information in $\hat{\rho}$ will understate DY ’s predictive power.

Empirically, using the information in $\hat{\rho}$ dramatically strengthens the case for predictability. When NYSE returns are regressed on log DY from 1946–2000, the least-squares (OLS) slope estimate is 0.92 with a standard error of 0.48. Stambaugh’s (1999) bias correction yields an estimate of 0.20 with a one-sided p -value of 0.308. However, using the information in $\hat{\rho}$, the bias-adjusted estimate becomes 0.66 with a t -statistic of 4.67, significant at the 0.000 level. (I stress that the estimate 0.66 is conservative, calculated under the assumption that $\rho \approx 1$; it is biased downward if ρ is truly less than one.) Predictability is also strong in various subsamples. For the first half of the sample, 1946–1972, the bias-adjusted estimate is 0.84 with a p -value less than 0.001. For the second half of the sample, 1973–2000, the bias-adjusted estimate is 0.64 with a p -value of 0.000. In short, by recognizing the upper bound on ρ , we obtain much stronger evidence of predictability.

As an aside, I also consider how the last few years of the sample affect the empirical results. DY reached a new low in May 1995, predicting that returns going forward should be far below average. In reality, the NYSE index more than doubled over the subsequent six years. When returns for 1995–2000 are added to the regression, the OLS slope coefficient drops in half, from 2.23 to 0.92, and the statistical significance declines from 0.068 to 0.308 using Stambaugh’s small-sample distribution. Interestingly, the tests here are relatively insensitive to the recent data. The bias-adjusted slope drops from 0.98 to 0.66 and the p -value remains 0.000. The reason is simple: the last few years have also lead to a sharp rise in the sample autocorrelation of DY , from 0.986 to 0.997. This rise means that the maximum bias

in the predictive slope declines from 1.25 to 0.25, offsetting most of the decline in the OLS estimate. Regressions with the equal-weighted index are even more remarkable, providing stronger evidence of predictability after observing the recent data.

My tests also suggest that B/M and E/P can predict market returns, though the evidence is less reliable. Regressions with these ratios begin in 1963 when Compustat data become available. I find that B/M and E/P forecast both equal- and value-weighted NYSE returns over the period 1963–1994, but they predict only the equal-weighted index once data for 1995–2000 are included. The evidence for both periods is much stronger than in previous studies. Kothari and Shanken (1997) and Pontiff and Schall (1998) conclude that B/M has little predictive power after 1960, and Lamont (1998) finds no evidence that E/P , by itself, predicts quarterly returns from 1947–1994.

I should emphasize that the sampling distribution studied by Stambaugh (1999), and used in many empirical studies, is generally appropriate for testing predictability. The test developed here is useful only when the predictive variable's sample autocorrelation is close to one (otherwise, high values of ρ are unlikely anyway, so the constraint $\rho < 1$ provides little information). Also, Stambaugh studies Bayesian tests in which he sometimes imposes the restriction that $\rho < 1$. He shows that doing so leads to stronger evidence of predictability. My contribution is to show that the constraint also improves inferences in a frequentist setting. In fact, as I discuss further below, the approach here is similar in some ways to Stambaugh's Bayesian analysis. The tests are identical if the Bayesian approach starts with a point prior that $\rho = 1$ (and no information about β); any other prior that places zero weight on $\rho > 1$ would produce stronger rejections of the null.

The paper proceeds as follows. Section 2 discusses the properties of predictive regressions and formalizes the statistical tests. Section 3 describes the data and Section 4 presents the main empirical results. Section 5 concludes.

2. Predictive regressions

Predictive regressions are common in the finance literature. They have been used to test whether past prices, financial ratios, interest rates, and a variety of other macroeconomic variables can forecast stock and bond returns. This section reviews the properties of predictive regressions, borrowing liberally from Stambaugh (1986, 1999).

2.1. Assumptions

The paper focuses on the regression

$$r_t = \alpha + \beta x_{t-1} + \varepsilon_t, \quad (3a)$$

where r_t is the return in month t and x_{t-1} is a predictive variable known at the beginning of the month. It is easy to show that β must be zero if expected returns are constant. In all cases discussed here, the alternative hypothesis is that $\beta > 0$, so we

will be concerned with one-sided tests. The predictive variable is assumed to follow a stationary AR1 process:

$$x_t = \phi + \rho x_{t-1} + \mu_t, \quad (3b)$$

where $\rho < 1$. Since an increase in price leads to a decrease in DY , the residuals in (3a) and (3b) are negatively correlated. It follows that ε_t is correlated with x_t in the predictive regression, violating one of the assumptions of OLS (which requires independence at all leads and lags). For simplicity, the variables are assumed to be normally distributed.

Before continuing, I should briefly discuss the stationarity assumption. The empirical tests rely heavily on the assumption that ρ is not greater than one. Statistically, the tests remain valid if $\rho = 1$ (we just need an upper bound on ρ), but I assume that ρ is strictly less than one to be consistent with prior studies. It also makes little sense to predict returns with a nonstationary variable. Economically, x_t should be stationary unless there is an explosive bubble in stock prices. Suppose, for example, that x_t equals $\log DY$. Then x_t is stationary if \log dividends and \log prices are cointegrated, implying that, in the long run, dividends and prices grow at the same rate. That assumption seems reasonable: there is a vast literature arguing against explosive bubbles and much evidence that DY is mean reverting over long sample periods.²

Of course, DY might exhibit other forms of nonstationarity that do not imply explosive behavior. For example, Fama and French (2002) suggest that the equity premium dropped sharply between 1951 and 2000. If the drop is permanent, not caused by transitory sentiment or the business cycle, it should lead to a permanent drop in DY . This type of nonstationarity could be modeled as a change in the intercept, ϕ , and regression (3b) can be thought of as a restricted version of a model with time-varying parameters. For my purposes, the empirical tests should be relatively insensitive to this type of nonstationarity as long as ρ remains below one. More important, if the decline in DY is really caused by a permanent drop in the risk premium, then we are already acknowledging that DY tracks changes in expected returns. Thus, this story does not say that we might falsely reject the null because it inherently assumes that the null is false.³

2.2. Properties of OLS

Denote the matrix of regressors as X , the coefficient vectors as $b = (\alpha \beta)'$ and $p = (\phi \rho)'$, and the residual vectors as ε and μ . The OLS estimates of Eqs. (3a)

²Blanchard and Watson (1982), Tirole (1982, 1985), and Loewenstein and Willard (2000) discuss the theory of rational bubbles; Hamilton and Whiteman (1985), Flood and Hodrick (1986, 1990), Diba and Grossman (1988), and West (1988) provide empirical evidence.

³Alternatively, a permanent shift in DY might occur because of a change in payout policy. But much of the recent decline in DY seems attributable to an increase in prices, not a decrease in dividends. Also, the empirical results are robust to leaving out the last 5, 10, or 15 years of data when repurchases grew in importance. The tests also use B/M and E/P , which are largely unaffected by payout policy.

and (3b) are then

$$\hat{b} = b + (X'X)^{-1}X'\varepsilon, \quad (4a)$$

$$\hat{p} = p + (X'X)^{-1}X'\mu. \quad (4b)$$

In the usual OLS setting, the estimation errors are expected to be zero. That is not true here: autocorrelations are biased downward in finite samples, and this bias feeds into the predictive regression through the correlation between ε_t and μ_t . Specifically, we can write $\varepsilon_t = \gamma\mu_t + v_t$, where $\gamma = \text{cov}(\varepsilon, \mu)/\text{var}(\mu)$. Substituting into (4a) yields

$$\hat{b} - b = \gamma(\hat{p} - p) + \eta, \quad (5)$$

where $\eta \equiv (X'X)^{-1}X'v$. The variable v_t is independent of μ_t , and consequently x_t , at all leads and lags. It follows that η has mean zero and variance $\sigma_v^2(X'X)^{-1}$.

Eq. (5) provides a convenient way to think about predictive regressions. Consider, first, the marginal distribution of $\hat{\beta}$ based on repeated sampling of both $\hat{\rho}$ and η . This distribution is studied by [Stambaugh \(1999\)](#) and commonly used for empirical tests. Eq. (5) implies that $\hat{\beta}$ inherits many properties of autocorrelations. For example, taking expectations yields

$$E[\hat{\beta} - \beta] = \gamma E[\hat{\rho} - \rho]. \quad (6)$$

Sample autocorrelations are biased downward by roughly $-(1 + 3\rho)/T$, inducing an upward bias in the predictive slope ($\gamma < 0$). Further, autocorrelations are negatively skewed and more variable than suggested by OLS. These properties imply that $\hat{\beta}$ is positively skewed and more variable than suggested by OLS. [Stambaugh](#) discusses these results in detail.

The tests below emphasize the conditional distribution of $\hat{\beta}$ given $\hat{\rho}$. Eq. (5) shows that, although $\hat{\beta}$ and $\hat{\rho}$ are not bivariate normal due to the irregularities in sample autocorrelations, $\hat{\beta}$ is normally distributed conditional on $\hat{\rho}$.⁴ This observation follows from the definition of η in Eq. (5). The conditional expectation of $\hat{\beta}$ is

$$E[\hat{\beta} - \beta | \hat{\rho}] = \gamma(\hat{\rho} - \rho), \quad (7)$$

which I refer to, informally, as the ‘realized bias’ in $\hat{\beta}$. (The conditional variance is given by $\text{var}(\eta)$, given above.) Eq. (7) implies that, if we knew $\hat{\rho} - \rho$, an unbiased estimator of β could be obtained by subtracting $\gamma(\hat{\rho} - \rho)$ from $\hat{\beta}$.

2.3. Statistical tests

The tests in this paper focus on the conditional distribution of $\hat{\beta}$. The idea is simple. Even though we don’t know $\hat{\rho} - \rho$, we can put a lower bound on it by assuming that $\rho \approx 1$. This assumption, in turn, gives an upper bound on the ‘bias’ in

⁴Technically, this statement requires that we condition on X , as in standard regression analysis, because it affects the variance of η ; see Eq. (5). If we condition only on $\hat{\rho}$, the distribution of β would be a mixture of normal distributions, all with mean zero but different variances.

$\hat{\beta}$. To be precise, define the bias-adjusted estimator

$$\hat{\beta}_{\text{adj}} = \hat{\beta} - \gamma(\hat{\rho} - \rho). \quad (8)$$

Given the true ρ , this estimator is normally distributed with mean β and variance $\sigma_v^2(X'X)_{(2,2)}^{-1}$. The autocorrelation is unknown, but as long as DY is stationary, the most conservative assumption for testing predictability is that $\rho \approx 1$: the bias in (7) is maximized, and the estimator in (8) is minimized, if we assume $\rho \approx 1$. If $\hat{\beta}_{\text{adj}}$ is significantly different from zero under this assumption, then it must be even more significant given the true value of ρ . (To implement the tests, we must also estimate γ and σ_v from $\varepsilon_t = \gamma\mu_t + v_t$. I use OLS estimates based on the sample values of ε_t and μ_t . The appendix shows how estimation error in γ and σ_v affects the statistical tests.)

Prior studies focus on the marginal distribution of $\hat{\beta}$, which implicitly assumes that we have no information about $\hat{\rho} - \rho$. That assumption is fine when $\hat{\rho}$ is small because the constraint $\rho < 1$ provides little information (high values of ρ are unlikely anyway). But the tests ignore useful information when $\hat{\rho}$ is close to one. Suppose, for example, that $\hat{\rho} = 0.99$ and $T = 300$. In this case, the bias in $\hat{\rho}$ is roughly $E[\hat{\rho} - \rho] = -0.016$. However, given the observed autocorrelation, the *minimum* possible value of $\hat{\rho} - \rho$ is actually -0.010 . The conditional test uses this information, together with the strong correlation between $\hat{\beta}$ and $\hat{\rho}$, to improve inferences about the predictive slope.

There is a second way to think about the test. The analysis above shows that sampling errors in $\hat{\beta}$ and $\hat{\rho}$ are closely related: $\hat{\beta}$ is expected to be high only when $\hat{\rho}$ is very low. Therefore, it is unlikely that we would observe both a high value of $\hat{\beta}$ and a high value of $\hat{\rho}$ (i.e., close to one) if the true parameters are $\beta = 0$ and $\rho < 1$. The empirical tests simply formalize this idea. They ask: under the null, what is the probability of observing such a high value for $\hat{\beta}$ given that $\hat{\rho}$ is so close to one? If we reject the joint hypothesis that $\beta = 0$ and $\rho < 1$, I interpret it as evidence that $\beta \neq 0$ because there are strong reasons, a priori, to believe that DY is stationary.

Fig. 1 illustrates these ideas. Panel A shows the marginal distribution of $\hat{\beta}$ and Panel B shows the joint distribution of $\hat{\beta}$ and $\hat{\rho}$. For the simulations, $\beta = 0$, $\rho = 0.99$, $T = 300$, and the correlation between ε_t and μ_t is -0.92 . Panel A clearly shows the strong bias and skewness in $\hat{\beta}$, with more than 85% of the estimates being positive. Panel B shows the strong correlation between $\hat{\beta}$ and $\hat{\rho}$. Comparing the two graphs, it is clear that high values of $\hat{\beta}$ correspond to samples in which $\hat{\rho}$ is far below ρ . Given that $\hat{\rho}$ is, say, no more than 0.01 below its true value, $\hat{\beta}$ is rarely greater than 0.40. That observation is key for the tests in this paper.

2.4. Using and combining the tests

The decision to use the conditional test or the standard ‘unconditional’ approach is, in some ways, quite simple. The unconditional approach is more general and gives the better estimate of β unless DY ’s autocorrelation is close to one. The conditional test, on the other hand, is useful when $\hat{\rho} - 1$ (the minimum value of $\hat{\rho} - \rho$) is greater than the expected bias in $\hat{\rho}$. With 25 years of data, this requires a monthly autocorrelation around 0.98 and an annual autocorrelation around 0.85; with 50 years of data, the values are 0.99 and 0.90, respectively. The conditional approach

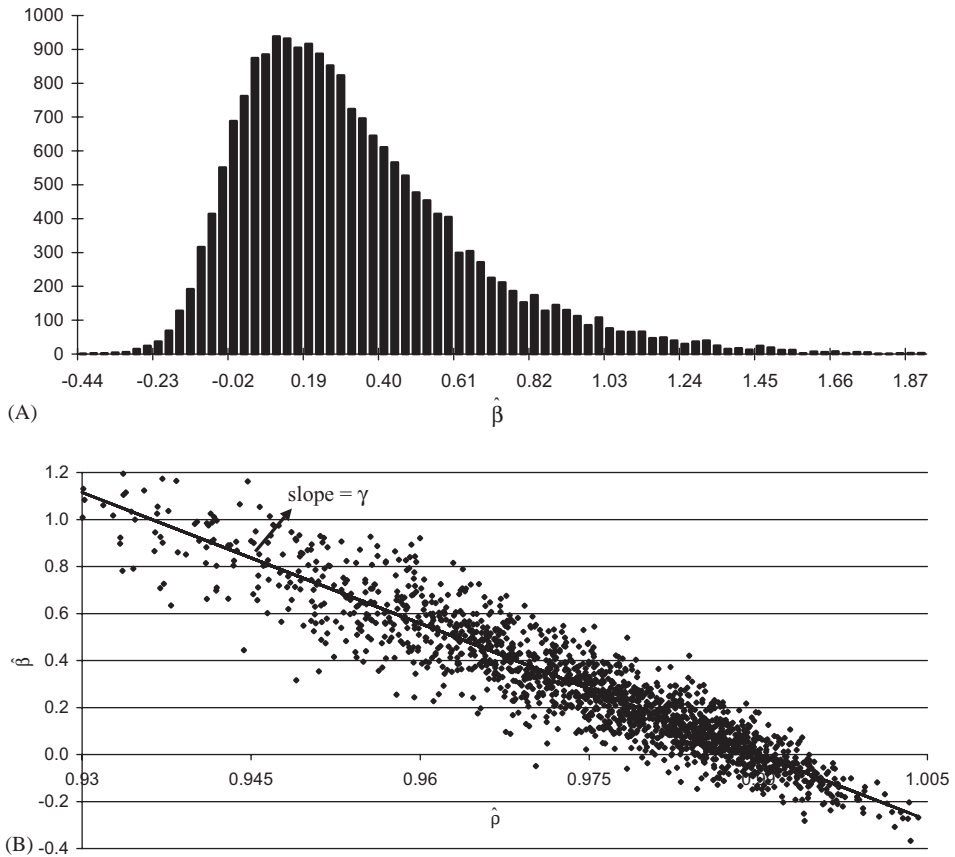


Fig. 1. Sampling distribution of $\hat{\beta}$ and $\hat{\rho}$. The figure shows the distribution of the OLS slope estimates from $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$ and $x_t = \phi + \rho x_{t-1} + \mu_t$. Panel A shows the marginal, or unconditional, distribution of $\hat{\beta}$ and Panel B shows the joint distribution of $\hat{\beta}$ and $\hat{\rho}$. The plots are based on Monte Carlo simulations (20,000 in Panel A and 2,000 in Panel B). The true parameters are $\beta = 0$, $\rho = 0.99$, $\text{cor}(\varepsilon, \mu) = -0.92$, $\sigma_\varepsilon = 0.04$, $\sigma_\mu = 0.002$, and $T = 300$. Panel A: Marginal distribution of $\hat{\beta}$. Panel B: Joint distribution of $\hat{\beta}$ and $\hat{\rho}$.

might still be useful if the autocorrelation is a bit lower, but it would depend on the underlying parameters (the empirical tests later illustrate this point).

The previous paragraph suggests that DY 's sample autocorrelation determines whether the conditional or unconditional test is better. Ideally, we could choose between the tests in advance, before looking at the data. From an ex ante perspective, the conditional test has greater power when ρ is close to one, but the opposite is true once ρ drops below some level that depends on the other parameters. (The appendix discusses power in more detail.) However, without prior information about ρ , we can't say ahead of time which test is better. Thus, it makes sense to rely on both tests and to calculate an overall significance level that reflects the probability of rejecting using either test.

The easiest way to calculate a joint significance level is simply to double the smaller of the two stand-alone p -values. The result is Bonferroni's upper bound on the true joint p -value.⁵ But notice that the conditional test, and consequently the Bonferroni upper bound, is very conservative if ρ is far from one; doubling the smaller p -value in this case—almost certainly from the unconditional test—isn't necessary. This suggests that a tighter upper bound might be derived using information about ρ . Intuitively, if $\hat{\rho}$ is, say, 0.50, the true ρ is likely to be far from one, and there is no reason to use either the conditional test or the Bonferroni p -value. This reasoning motivates the following joint test, a modification to Bonferroni:

An overall p -value, when the conditional and unconditional tests are both used, is given by $\min(2P, P + D)$, where P is the smaller of the two stand-alone p -values and D is the p -value for testing $\rho = 1$, based on the sampling distribution of $\hat{\rho}$.

The first part, $2P$, is just the Bonferroni upper bound; the second part, $P + D$, recognizes that doubling P is too conservative if we reject $\rho = 1$ (that is, if D is small). The bound implies that if, ex post, the autocorrelation appears to be far from one ($D \approx 0$), we can just use the unconditional test's p -value as the overall p -value, without doubling. The adjustment for searching becomes larger if $\hat{\rho}$ is close to one, with a maximum provided by Bonferroni.

It is difficult to prove that the bound holds for all ρ but simulations suggest that it does. The bound can be shown to hold if ρ is either very far from or very close to one.⁶ It should hold for intermediate values of ρ as well, but a rigorous proof is difficult because the sampling distribution of $\hat{\beta}$ and $\hat{\rho}$ is analytically intractable. To test the bound, I ran simulations calibrated to the full-sample and first half-sample DY regressions, allowing ρ to vary from 0.9 to 0.9999. I calculated $\min(2P, P + D)$ for each simulation and rejected the null if the value was below 0.05. The rejection rate was less than or equal to 5% for all ρ . The rejection rates seem to be U-shaped: they are close to 5% for $\rho = 0.9999$ and $\rho < 0.96$, but less than 5% for intermediate values.

In sum, the conditional test is a natural addition to the small-sample distribution used by prior studies. Because we cannot say, ex ante, whether the conditional or unconditional approach is better, a simple algorithm is to use both tests and calculate a joint significance level. The modified Bonferroni approach provides a way to calculate the joint p -value.

⁵Suppose we run two tests and reject if either A or B occurs. Each has probability P under the null. The overall probability of rejecting is $\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \text{ and } B) < 2P$.

⁶The bound is derived from asking: Under the null, what is the probability that either test rejects with p -value P and we reject with p -value D that $\rho = 1$? Let A , B , and C be the rejection regions for the conditional test, unconditional test, and unit root test, respectively. $\Pr((A \text{ or } B) \text{ and } C) = \Pr((A \text{ and } C) \text{ or } (B \text{ and } C)) < \Pr(A) + \min(\Pr(B), \Pr(C))$. If $\rho = 1$, the bound obtains because $\Pr(A) = \Pr(B) = P$ and $\Pr(C) = D$. As ρ drops, $\Pr(C)$ increases but $\Pr(A)$ decreases, and the bound holds as long as the second effect dominates. When ρ is small, $\min(2P, P + D)$ converges, appropriately, to the unconditional test's p -value.

The appendix discusses power in more detail. It shows that combining the conditional and unconditional tests has advantages over using either test individually. In particular, a joint test always sacrifices power to one of the individual tests in specific regions of the parameter space, but performs best globally over a large range of possible parameters.

3. Data and descriptive statistics

I use the above methodology to test whether DY , B/M , and E/P can forecast market returns. Prices and dividends come from the Center for Research in Security Prices (CRSP) database. Earnings and book values come from Compustat. The tests focus on NYSE equal- and value-weighted indices to be consistent with prior research and to avoid changes in the market's composition as AMEX and NASDAQ firms enter the database.

DY is calculated monthly on the value-weighted NYSE index. It is defined as dividends paid over the prior year divided by the current level of the index. Thus, DY is based on a rolling window of annual dividends. I use value-weighted DY to predict returns on both the equal- and value-weighted indices. Value-weighted DY is likely to be a better measure of aggregate dividend yield (it equals total dividends divided by total market value). The predictive regressions use the natural log of DY , rather than the raw series, because it should have better time-series properties. Raw DY , measured as a ratio, is likely to be positively skewed and its volatility depends mechanically on its level (when DY is two, price movements must be twice as big to have the same effect as when DY is four). Taking logs solves both of these problems.

The empirical tests with DY focus on the period January 1946 to December 2000. I omit the Depression era because the properties of stock prices were much different prior to 1945. Returns were extremely volatile in the 1930s, and this volatility is reflected in both the variance and persistence of DY (see Fama and French, 1988). As a robustness check, I split the sample in half and look at the two subperiods, 1946–1972 and 1973–2000. Further, I investigate the influence of the last few years because recent stock returns have been so unusual.

The tests with B/M and E/P are restricted to the Compustat era, 1963–2000. B/M is the ratio of book equity for the previous fiscal year to market equity in the previous month. Similarly, E/P is the ratio of operating earnings (before depreciation) in the previous fiscal year to market equity in the previous month. I use operating earnings because Shiller (1984) and Fama and French (1988) suggest that net income is a noisy measure of fundamentals; preliminary tests suggest that operating earnings are a better measure.⁷ To ensure that the tests are predictive, I do

⁷Log E/P ratios are highly autocorrelated using either measure: 0.990 for operating earnings and 0.989 for net income. However, the residuals in an AR1 regression are more variable when E/P is based on net income (standard deviation of 0.062 compared with 0.049 for operating earnings) and not as highly correlated with returns (−0.66 vs. −0.86). These observations might indicate more noise in the measurement of net income. In any case, the predictive power of the two series is similar and, for simplicity, I report only tests using operating earnings (the results for net income are marginally weaker).

not update accounting numbers until four months after the fiscal year. Also, to reduce possible selection biases, a firm must have three years of accounting data before it is included in the sample (see Kothari et al., 1995). The regressions use $\log B/M$ and $\log E/P$, both measured on the value-weighted NYSE index.

Table 1 provides summary statistics for the data. Volatility is bit lower in the first half of the sample, but most of the time-series properties are stable across the two subperiods. DY averages 3.80% over the full sample with a standard deviation of 1.20%. Since DY is a ratio, $\log DY$ should better approximate a normal distribution. The table confirms that $\log DY$ is more symmetric in the first half of the sample, but it is negatively skewed in the second half (primarily due to 1995–2000). The properties of B/M and E/P are similar to those of DY . B/M averages 0.53 and E/P

Table 1
Summary statistics, 1946–2000

The table reports summary statistics for stock returns, dividend yield, book-to-market, and the earnings-price ratio. Observations are monthly and the variables are expressed in percent; $\log(x)$ equals the natural log of x expressed in percent. Prices and dividends come from CRSP and accounting data come from Compustat. EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. The financial ratios are all calculated for the value-weighted index. DY equals dividends paid over the prior year divided by the current level of the index; B/M is the ratio of book equity to market equity; E/P is the ratio of operating earnings before depreciation to market equity.

Variable	Mean	S.D.	Skew.	Autocorrelation		
				ρ_1	ρ_{12}	ρ_{24}
<i>Returns and dividend yield</i>						
Full sample: 1946–2000						
VWNY	1.04	4.08	−0.38	0.032	0.042	0.014
EWNY	1.11	4.80	−0.16	0.136	0.065	0.027
DY	3.80	1.20	0.37	0.992	0.889	0.812
$\log(DY)$	1.28	0.33	−0.53	0.997	0.948	0.912
1st half: 1946–1972						
VWNY	0.98	3.67	−0.39	0.079	0.026	0.066
EWNY	1.04	4.38	−0.26	0.150	0.024	0.021
DY	4.02	1.21	0.84	0.992	0.879	0.774
$\log(DY)$	1.35	0.28	0.56	0.993	0.876	0.785
2nd half: 1973–2000						
VWNY	1.10	4.44	−0.39	0.001	0.065	−0.013
EWNY	1.18	5.18	−0.11	0.125	0.113	0.030
DY	3.59	1.15	−0.19	0.991	0.899	0.914
$\log(DY)$	1.22	0.37	−0.81	0.999	0.996	1.062
<i>Book-to-market and earnings-price ratio</i>						
Compustat: 1963–2000						
B/M	53.13	18.28	0.39	0.990	0.891	0.837
$\log(B/M)$	3.91	0.36	−0.19	0.995	0.951	0.923
E/P	20.02	7.01	0.55	0.988	0.864	0.770
$\log(E/P)$	2.94	0.35	0.14	0.990	0.891	0.785

averages 0.20. The raw series for B/M and E/P are positively skewed, while the log series are nearly symmetric. (The E/P ratio appears high because earnings are operating earnings before depreciation. The E/P ratio based on net income averages 0.07 over the period.)

Table 1 also shows that the financial ratios are extremely persistent. The first-order autocorrelations range from 0.988 to 0.999 for the various series. The autocorrelations tend to diminish as the lag increases, but DY in the second half of the sample is essentially a random walk. Log DY , B/M , and E/P are more highly autocorrelated than the raw series. This is important for the empirical tests since the bias-adjustment depends on $\hat{\rho} - 1$.

4. Empirical results

This section tests whether the financial ratios predict market returns. I initially focus on DY because it has been studied most in the literature, and I pay special attention to the last few years of the sample because they have a large impact on the results.

4.1. Predicting with DY

Table 2 explores the predictive power of DY in the full sample, 1946–2000. The table estimates the model analyzed in Section 2:

$$r_t = \alpha + \beta x_{t-1} + \varepsilon_t, \quad (9a)$$

$$x_t = \phi + \rho x_{t-1} + \mu_t, \quad (9b)$$

where r_t is the stock return and x_{t-1} is the dividend yield. I estimate regressions for NYSE equal- and value-weighted returns and for nominal and excess returns (measured net of the one-month T-bill rate). All of the regressions use DY for the value-weighted NYSE index.

The table reports a variety of statistics. The row labeled ‘OLS’ shows the least-squares slope and standard error. These estimates ignore bias and are reported primarily as a benchmark. The row labeled ‘Stambaugh’ shows estimates based on Stambaugh’s (1999) results. The point estimate, standard error, and p -value are all adjusted for bias using the marginal distribution of $\hat{\beta}$, obtained from Monte Carlo simulations. The distribution depends on the unknown parameters ρ and Σ , for which I substitute the OLS estimates. Stambaugh notes that the distribution is insensitive to small changes in the parameters, so this substitution should not be too important (Campbell and Yogo, 2003, provide a more rigorous way to account for uncertainty about ρ).

The final row, labeled ‘ $\rho \approx 1$,’ reports estimates based on the conditional distribution of $\hat{\beta}$. The slope coefficient is the bias-adjusted estimator

$$\hat{\beta}_{\text{adj}} = \hat{\beta} - \gamma(\hat{\rho} - \rho), \quad (10)$$

Table 2

Dividend yield and expected returns, 1946–2000

The table reports an AR1 regression for dividend yield and predictive regressions for stock returns for the period January 1946–December 2000 (660 months). Observations are monthly. DY is the dividend yield on the value-weighted NYSE index and $\text{Log}(DY)$ is the natural logarithm of DY . EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns are calculated as EWNY and VWNY minus the one-month T-bill rate. All data come from CRSP; returns are expressed in percent. For the predictive regressions, ‘OLS’ reports the standard OLS estimates, ‘Stambaugh’ reports the bias-adjusted estimate and p -value based on Stambaugh (1999), and ‘ $\rho \approx 1$ ’ reports the bias-adjusted estimate and p -value assuming that ρ is approximately one.

$\text{Log}(DY_t) = \phi + \rho \text{Log}(DY_{t-1}) + \mu_t$		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
AR(1)	OLS	0.997	0.005	−0.008	−0.006	0.984	0.043
$r_t = \alpha + \beta \text{Log}(DY_{t-1}) + \varepsilon_t$		β	S.E.(β)	p -value	Adj. R^2	S.D.(ε)	cor(ε, μ)
VWNY	OLS	0.917	0.476	0.027	0.004	4.068	−0.955
	Stambaugh	0.196	0.670	0.308			
	$\rho \approx 1$	0.663	0.142	0.000			
EWNY	OLS	1.388	0.558	0.007	0.008	4.773	−0.878
	Stambaugh	0.615	0.758	0.183			
	$\rho \approx 1$	1.115	0.268	0.000			
Excess VWNY	OLS	0.915	0.478	0.028	0.004	4.087	−0.953
	Stambaugh	0.192	0.673	0.311			
	$\rho \approx 1$	0.661	0.145	0.000			
Excess EWNY	OLS	1.387	0.560	0.007	0.008	4.788	−0.878
	Stambaugh	0.611	0.760	0.184			
	$\rho \approx 1$	1.112	0.268	0.000			

where I assume that ρ is approximately one (operationalized as $\rho = 0.9999$). If ρ is truly less than one, this estimate is biased downward and the test understates the predictive power of DY . The variance of $\hat{\beta}_{\text{adj}}$ is $\sigma_v^2(X'X)_{(2,2)}^{-1}$. To implement the test, we need to estimate γ and σ_v from $\varepsilon_t = \gamma\mu_t + v_t$, where ε_t and μ_t are the residuals in (9). The appendix describes how estimation error in these parameters affects the results. Under the null, the resulting t -statistic is truly a t -statistic—that is, it has a Student t distribution with $T - 3$ degrees of freedom.

Table 2 provides strong evidence of predictability. Consider, first, the nominal return on the value-weighted index (VWNY). The OLS slope is 0.92 with a standard error of 0.48. The point estimate implies that a one-standard-deviation change in DY (equal to 0.33) predicts a 0.30% change in monthly expected return. Based on Stambaugh’s (1999) distribution, the bias-adjusted estimate shrinks to 0.20 with an unimpressive p -value of 0.308. The conditional test, however, gives a much different picture. Assuming that $\rho < 1$, the ‘realized bias’ in the predictive slope is *at most* 0.25,

implying a bias-adjusted estimate of 0.66 with a t -statistic of 4.67 (p -value of 0.000). The strong significance in the conditional test is, in part, attributable to a standard error that is much lower than the standard error from Stambaugh's distribution, 0.14 vs. 0.67. (Note that the standard error is relevant only for testing a lower bound on the predictive slope, i.e., for testing the null of no predictability; if DY 's autocorrelation is truly less than one, β might be substantially larger than suggested by a symmetric confidence interval around the conditional estimate.)

The regressions for equal-weighted NYSE returns (EWNY) confirm these findings. The conditional bias-adjusted slope, 1.12, is larger than the estimate for VWNY and suggests significant time-variation in expected returns (p -value of 0.000). The table also shows that nominal and excess returns produce very similar estimates during this period.

The results just described focus on stand-alone p -values for the conditional and unconditional (Stambaugh) tests. As discussed earlier, an overall significance level should recognize that we 'search' over multiple tests. For the regressions in Table 2, the joint p -value suggested in Section 2 amounts to a simple Bonferroni correction, which doubles the smaller stand-alone p -value. This joint test obviously rejects quite strongly since the p -values from the conditional tests are so close to zero; in fact, the p -values all remain 0.000 after doubling.

These results show that the small-sample distribution analyzed by Stambaugh (1986, 1999) and Mankiw and Shapiro (1986) can greatly understate the significance of DY . Their tests do not make use of the constraint that $\rho < 1$, which lowers the point estimate and raises the standard error when $\hat{\rho}$ is close to one. Surprisingly, the conditional tests in Table 2 show that DY is more significant than suggested by OLS. The bias-adjusted slopes are lower than the OLS estimates, but the information conveyed by $\hat{\rho}$ has an even bigger effect on the standard errors. As a result, the stochastic properties of DY actually strengthen the case for predictability.

I should mention that conditioning on $\hat{\rho}$ would not have helped in Stambaugh's (1999) tests, primarily because DY is not as highly autocorrelated in his sample. There are two reasons: (1) Stambaugh uses raw DY , which is slightly less persistent than $\log DY$, and (2) DY is not as highly autocorrelated during his sample periods (e.g., 1926–1996). I have repeated the tests in Table 2 using Stambaugh's post-war sample of 1952–1996 and continue to find significant predictability in the conditional tests. Regressions with nominal returns are similar to those in Table 2, while estimates for excess returns drop considerably, in part because the correlation between DY and T-bill rates is sensitive to the starting date of the sample.⁸

Table 3 reports results for the first and second halves of the sample, 1946–1972 and 1973–2000. The tests strongly reject the null in most cases even though the periods are quite short. For 1946–1972, DY predicts excess returns on both indices. The bias-adjusted slope for excess VWNY, 1.16 with a p -value of 0.000, is larger than

⁸The correlation between T-bill and DY is 0.00 for 1946–2000 and 0.24 for 1952–2000. Regressions with raw DY and excess returns are especially sensitive to dropping the years 1946–1951; raw DY has strong predictive power from 1946–2000 but not from 1952–2000. The results are more stable using nominal returns, or if the regressions control directly for T-bill rates (details available on request).

Table 3

Dividend yield and expected returns, 1946–1972 and 1973–2000

The table reports ARI regressions for dividend yield and predictive regressions for stock returns for two periods, January 1946–December 1972 and January 1973–December 2000 (324 and 336 months, respectively). Observations are monthly. DY is the dividend yield on the value-weighted NYSE index and $\text{Log}(DY)$ is the natural logarithm of DY . EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns are calculated as EWNY and VWNY minus the one-month T-bill rate. All data come from CRSP; returns are expressed in percent. For the predictive regressions, ‘OLS’ reports the standard OLS estimates, ‘Stambaugh’ reports the bias-adjusted estimate and p -value based on Stambaugh (1999), and ‘ $\rho \approx 1$ ’ reports the bias-adjusted estimate and p -value assuming that ρ is approximately one.

		1946–1972						1973–2000					
		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)	ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
Log(DY_t) = $\phi + \rho \text{Log}(DY_{t-1}) + \mu_t$													
AR(1)	OLS	0.993	0.008	-0.016	-0.012	0.981	0.039	0.999	0.007	-0.016	-0.012	0.984	0.046
Log(DY_t) = $\phi + \rho \text{Log}(DY_{t-1}) + \mu_t$													
		1946–1972						1973–2000					
		β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)	β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)
VWNY	OLS	1.421	0.723	0.025	0.009	3.649	-0.935	0.732	0.665	0.136	0.001	4.441	-0.968
	Stambaugh	0.013	1.400	0.405				-0.751	1.254	0.700			
	$\rho \approx 1$	0.844	0.257	0.001				0.641	0.167	0.000			
EWNY	OLS	1.077	0.863	0.107	0.002	4.359	-0.884	1.703	0.770	0.014	0.012	5.142	-0.875
	Stambaugh	-0.511	1.637	0.551				0.152	1.397	0.367			
	$\rho \approx 1$	0.425	0.405	0.152				1.608	0.373	0.000			
Exc.VWNY	OLS	1.733	0.724	0.009	0.014	3.656	-0.933	0.395	0.669	0.278	-0.002	4.467	-0.968
	Stambaugh	0.325	1.402	0.327				-1.096	1.261	0.835			
	$\rho \approx 1$	1.156	0.261	0.000				0.304	0.169	0.041			
Exc.EWNY	OLS	1.388	0.864	0.055	0.005	4.364	-0.883	1.366	0.774	0.039	0.006	5.174	-0.875
	Stambaugh	-0.199	1.639	0.460				-0.194	1.406	0.474			
	$\rho \approx 1$	0.736	0.407	0.037				1.271	0.376	0.000			

in the full sample, while the slope for excess EWNV is somewhat smaller, 0.74 with a p -value of 0.037. In the second half of the sample, DY predicts raw and excess returns on both indices with p -values between 0.000 and 0.041. Estimates for the value-weighted index drop from the earlier period, but estimates for the equal-weighted index are especially large. The bias-adjusted slope for excess EWNV, 1.27, implies that a one-standard-deviation change in DY is associated with a 0.47% change in monthly expected return.

The subperiod results reveal striking differences between the conditional and unconditional tests. Consider the VWNV regressions in the second half of the sample. The bias-adjusted estimate from the unconditional tests is -0.75 with a standard error of 1.25. In contrast, the estimate from the conditional tests is 0.64 with a standard error of 0.17. The p -values for the two tests are 0.700 and 0.000, respectively. Thus, incorporating the information in $\hat{\rho}$ can be critical when the autocorrelation is close to one and the sample is relatively short. The AR1 regressions for DY , at the top of the table, show why. The expected bias in $\hat{\rho}$ is approximately -0.016 , while the minimum realized value of $\hat{\rho} - \rho$ is -0.001 . As a consequence, the maximum ‘realized bias’ in the predictive slope is more than 90% smaller than the unconditional estimate.

I also emphasize that the conditional tests are quite conservative: the conditional estimate of β is biased downward if ρ is truly less than one. To illustrate this point, it is useful to explore how alternative assumptions about ρ affect the results. Consider the regression for nominal EWNV for 1946–1972. The p -value is 0.147 if we assume that $\rho \approx 1$ but it drops to 0.004 if we assume that $\rho = 0.993$ (the sample autocorrelation). The p -value is less than 0.050 for any $\rho < 0.997$. While I am not suggesting that DY is truly significant in this regression, since it is difficult to justify any particular value for ρ less than one, the reported p -values do appear to be quite conservative.

Bayesian methods provide a more rigorous way to incorporate uncertainty about ρ (see also [Stambaugh, 1999](#)). Suppose an investor begins with no information about β . It is straightforward to show, using standard results for this model, that the Bayesian posterior probability for $\beta \leq 0$ can be found by integrating the conditional p -value (conditional on a given ρ) over the posterior distribution of ρ . Consider three different beliefs about ρ . If the investor believes $\rho \approx 1$ with certainty, the posterior probability for $\beta \leq 0$ just equals the conditional p -value in [Table 3](#). This p -value is 0.147 for nominal EWNV. If, instead, the investor begins with a flat prior for $\rho \leq 1$, the posterior probability would drop to 0.017. Finally, suppose that we arbitrarily shift the investor’s posterior belief about ρ upward by one standard deviation and truncate at one. This posterior represents fairly strong beliefs that ρ is close to one; it is roughly the lower tail of a normal distribution with mean one and standard deviation 0.008 (the standard error of $\hat{\rho}$). In this case, the posterior probability for $\beta \leq 0$ equals 0.032. Once again, the p -values in [Table 3](#) seem quite conservative.

4.2. The influence of 1995–2000

The tests in [Tables 2 and 3](#) include data for 1995–2000, during which time prices moved strongly against the predictions of the model: DY was extremely low,

Table 4
Influence of 1995–2000

The table reports AR1 regressions for dividend yield and predictive regressions for stock returns for two periods, January 1946–December 1994 (588 months) and January 1946–December 2000 (660 months). Observations are monthly. *DY* is the dividend yield on the value-weighted NYSE index and $\text{Log}(DY)$ is the natural logarithm of *DY*. EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. All data come from CRSP; returns are expressed in percent. For the predictive regressions, ‘OLS’ reports the standard OLS estimates, ‘Stambaugh’ reports the bias-adjusted estimate and *p*-value based on Stambaugh (1999), and ‘ $\rho \approx 1$ ’ reports the bias-adjusted estimate and *p*-value assuming that ρ is approximately one.

		January 1946–December 1994					
$\text{Log}(DY_t) = \phi + \rho \text{Log}(DY_{t-1}) + \mu_t$		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
AR(1)	OLS	0.986	0.007	–0.008	–0.007	0.971	0.043
$r_t = \alpha + \beta \text{Log}(DY_{t-1}) + \varepsilon_t$		β	S.E.(β)	<i>p</i> -value	Adj. R^2	S.D.(ε)	cor(ε, μ)
VWNY	OLS	2.230	0.670	0.000	0.017	4.056	–0.952
	Stambaugh	1.532	0.939	0.068			
	$\rho \approx 1$	0.980	0.205	0.000			
EWNY	OLS	2.422	0.805	0.001	0.014	4.874	–0.882
	Stambaugh	1.645	1.117	0.081			
	$\rho \approx 1$	1.031	0.381	0.004			
		January 1946–December 2000					
$\text{Log}(DY_t) = \phi + \rho \text{Log}(DY_{t-1}) + \mu_t$		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
AR(1)	OLS	0.997	0.005	–0.008	–0.006	0.984	0.043
$r_t = \alpha + \beta \text{Log}(DY_{t-1}) + \varepsilon_t$		β	S.E.(β)	<i>p</i> -value	Adj. R^2	S.D.(ε)	cor(ε, μ)
VWNY	OLS	0.917	0.476	0.027	0.004	4.068	–0.955
	Stambaugh	0.196	0.670	0.308			
	$\rho \approx 1$	0.663	0.142	0.000			
EWNY	OLS	1.388	0.558	0.007	0.008	4.773	–0.878
	Stambaugh	0.615	0.758	0.183			
	$\rho \approx 1$	1.115	0.268	0.000			

dropping from 2.9% in January 1995 to 1.5% in December 2000, yet the value-weighted index nearly doubled. Because this period is so unusual, I briefly consider its effect on the results.

Table 4 shows estimates for 1946–1994 together with corresponding estimates for the full sample (the same as Table 2). I report only regressions with nominal returns for simplicity. Focusing on the value-weighted index, the OLS slope coefficient in the truncated sample is more than double its value for the full period, 2.23 compared

with 0.92. Using Stambaugh's (1999) small-sample distribution, the p -value is 0.068 in the truncated sample but 0.308 in the full regression. Interestingly, however, the conditional tests are much less sensitive to the recent data. The bias-adjusted slope drops from 0.98 to 0.66 and remains significant at the 0.000 level. The t -statistic is almost unchanged, declining from 4.80 to 4.68. Thus, the statistical significance of DY remains strong even though the model performed terribly after 1995.

The relative insensitivity of the conditional tests can be explained by the sample autocorrelation of DY . Table 4 shows that the sample autocorrelation increases from 0.986 to 0.997. This increase means that the sampling error in ρ must have gone up (become less negative or more positive). The conditional tests implicitly recognize that the sampling error in β has correspondingly decreased; the estimated bias in the slope, $\gamma(\hat{\rho} - 1)$, declines from 1.25 to 0.25. Although the OLS estimate drops by 1.31, the conditional test attributes 76% (1.00/1.31) of the decline to changes in $\hat{\rho}$.

Table 4 reports an even more remarkable result: for EWNY, the bias-adjusted slope actually increases with the addition of 1995–2000. Again, this counterintuitive result can be explained by the sharp rise in the sample autocorrelation of DY . Given the large drop in DY from 1995 to 2000, and the associated increase in $\hat{\rho}$, we would expect to see contemporaneously high returns and a decrease in the predictive slope. The conditional tests implicitly recognize this fact by adjusting the maximum bias for changes in $\hat{\rho} - 1$.

The truncated sample is interesting for another reason: the unconditional bias-adjusted slopes are higher than the conditional estimates but their significance levels are much lower. Focusing on value-weighted returns, the unconditional bias-adjusted slope is 1.53 with a p -value of 0.068; the conditional bias-adjusted slope is 0.98 with a p -value of 0.000. This combination is a bit awkward. The higher unconditional slope suggests that the conditional tests are too conservative; the true autocorrelation is probably lower than one. But, without the extreme assumption, we do not reject as strongly. This finding points to an odd property of the tests.

4.3. Predicting with B/M and E/P

Tables 5 and 6 extend the tests to B/M and E/P . Prior studies have used these ratios to predict market returns with limited success. I report regressions for the full sample, 1963–2000, and for the truncated sample ending in 1994.

B/M has some forecasting ability, but the evidence is less reliable than for DY . In the truncated sample, B/M is significant only for nominal returns. The bias-adjusted slopes are 1.11 and 0.73 for equal- and value-weighted returns, respectively, with (stand-alone) p -values of 0.022 and 0.017. Adding data for 1995–2000, B/M predicts both raw and excess returns on the equal-weighted index but has little power to predict VWNY. The estimates for EWNY are similar to those for the truncated sample, 1.03 for raw returns (p -value of 0.005) and 0.68 for excess returns (p -value of 0.047).

Like the earlier results, Table 5 shows that conditional tests provide much stronger evidence of predictability than unconditional tests. The differences are dramatic in the full sample. By incorporating the information in $\hat{\rho}$, the bias declines by more than

Table 5

Book-to-market and expected returns, 1963–1994 and 1963–2000

The table reports AR1 regressions for the book-to-market ratio and predictive regressions for stock returns for two periods, June 1963–December 1994 (379 months) and June 1963–December 2000 (451 months). Observations are monthly. B/M is the ratio of book equity to market equity on the value-weighted NYSE index. $\text{Log}(B/M)$ is the natural logarithm of B/M . EWNV and VWNV are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns equal EWNV and VWNV minus the one-month T-bill rate. Prices and returns come from CRSP, book equity comes from Compustat, and returns are expressed in percent. For the predictive regressions, ‘OLS’ reports standard OLS estimates, ‘Stambaugh’ reports the bias-adjusted estimate and p -value based on Stambaugh (1999), and ‘ $\rho \approx 1$ ’ reports the bias-adjusted estimate and p -value assuming that ρ is approximately one.

		1963–1994						1963–2000					
		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)	ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
$\text{Log}(B/M_t) = \phi + \rho \text{Log}(B/M_{t-1}) + \mu_t$													
AR(1)	OLS	0.987	0.009	-0.013	-0.010	0.972	0.047	0.995	0.006	-0.011	-0.009	0.983	0.047
		1963–1994						1963–2000					
		β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)	β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)
$r_t = \alpha + \beta \text{Log}(B/M_{t-1}) + \epsilon_t$													
VWNY	OLS	1.801	0.784	0.011	0.011	4.265	-0.901	0.698	0.594	0.108	0.001	4.237	-0.890
	Stambaugh	0.772	1.240	0.220				-0.222	0.945	0.515			
	$\rho \approx 1$	0.731	0.341	0.017				0.276	0.258	0.149			
EWNV	OLS	2.312	0.963	0.008	0.012	5.239	-0.826	1.484	0.670	0.014	0.009	5.032	-0.802
	Stambaugh	1.150	1.493	0.189				0.501	1.089	0.266			
	$\rho \approx 1$	1.107	0.545	0.022				1.032	0.400	0.005			
Exc.VWNY	OLS	1.275	0.790	0.054	0.004	4.298	-0.902	0.351	0.567	0.268	-0.001	4.260	-0.892
	Stambaugh	0.236	1.250	0.352				-0.576	0.950	0.698			
	$\rho \approx 1$	0.196	0.342	0.291				-0.075	0.257	0.626			
Exc.EWNV	OLS	1.786	0.969	0.033	0.006	5.274	-0.827	1.136	0.674	0.046	0.004	5.060	-0.803
	Stambaugh	0.615	1.503	0.287				0.147	1.096	0.367			
	$\rho \approx 1$	0.571	0.547	0.151				0.681	0.402	0.047			

Table 6

Earnings-price ratio and expected returns, 1963–1994 and 1963–2000

The table reports AR1 regressions for the earnings-price ratio and predictive regressions for stock returns for two periods, June 1963–December 1994 (379 months) and June 1963–December 2000 (451 months). Observations are monthly. E/P is the ratio of operating earnings to market equity on the value-weighted NYSE index. $\text{Log}(E/P)$ is the natural logarithm of E/P . EWNV and VWNV are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns equal EWNV and VWNV minus the one-month T-bill rate. Prices and returns come from CRSP, earnings come from Compustat, and returns are expressed in percent. For the predictive regressions, ‘OLS’ reports standard OLS estimates, ‘Stambaugh’ reports the bias-adjusted estimate and p -value based on Stambaugh (1999), and ‘ $\rho \approx 1$ ’ reports the bias-adjusted estimate and p -value assuming that ρ is approximately one.

		1963–1994						1963–2000					
		ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)	ρ	S.E.(ρ)	Bias	$-(1 + 3\rho)/T$	Adj. R^2	S.D.(μ)
AR(1)	OLS	0.987	0.008	-0.011	-0.010	0.978	0.049	0.990	0.007	-0.010	-0.009	0.980	0.049
		1963–1994						1963–2000					
		β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)	β	S.E.(β)	p -value	Adj. R^2	S.D.(ϵ)	cor(ϵ, μ)
VWNY	OLS	1.566	0.660	0.009	0.012	4.263	-0.866	1.121	0.576	0.026	0.006	4.226	-0.861
	Stambaugh	0.495	1.071	0.213				0.205	0.916	0.275			
	$\rho \approx 1$	0.566	0.332	0.046				0.403	0.294	0.088			
EWNV	OLS	1.950	0.811	0.008	0.012	5.239	-0.796	1.753	0.685	0.005	0.012	5.023	-0.778
	Stambaugh	0.655	1.295	0.194				0.691	1.062	0.159			
	$\rho \approx 1$	0.820	0.493	0.049				0.983	0.432	0.012			
Exc.VWNY	OLS	1.119	0.665	0.047	0.005	4.296	-0.868	0.725	0.580	0.106	0.001	4.255	-0.863
	Stambaugh	0.039	1.079	0.332				-0.198	0.923	0.424			
	$\rho \approx 1$	0.108	0.332	0.380				0.000	0.294	0.510			
Exc.EWNV	OLS	1.503	0.817	0.033	0.005	5.275	-0.798	1.358	0.689	0.025	0.006	5.054	-0.780
	Stambaugh	0.198	1.305	0.282				0.288	1.069	0.247			
	$\rho \approx 1$	0.363	0.495	0.236				0.580	0.433	0.093			

50% and the p -values for VWNY and EWNY drop from 0.515 and 0.266 to 0.142 and 0.005, respectively. The two tests are also much different comparing the full and truncated samples. The regressions for excess EWNY are especially interesting. The unconditional bias-adjusted slope is 0.62 for the sample ending in 1994 and 0.15 for the sample ending in 2000. But the conditional bias-adjusted slope actually *increases* slightly, from 0.57 to 0.68. That result mirrors the evidence for DY in Table 4.

The results for E/P in Table 6 are similar to those for B/M . E/P appears to forecast nominal returns, but there is little evidence that it forecasts excess returns. The p -values for nominal returns range from 0.012 to 0.088 for the different time periods and stock returns. In the full sample, a one-standard-deviation increase in E/P maps into a 0.14% increase in expected return for VWNY and a 0.34% increase for EWNY. Table 6 also confirms the influence of 1995–2000 on the regressions. The OLS slopes decline when the recent data are included, but the conditional bias-adjusted slopes remain approximately the same. For both nominal and excess EWNY, the addition of 1995–2000 again strengthens the case for predictability.

5. Summary and conclusions

The literature on stock predictability has evolved considerably over the last 20 years. Initial tests produced strong evidence that market returns are predictable, but subsequent research has argued that small-sample biases explain the bulk of apparent predictability. The accumulated evidence suggests that DY , B/M , and E/P have, at best, weak power to predict returns. In this paper, I provide a new test of predictability, emphasizing four main points:

(a) Previous studies focus on the marginal distribution of the predictive slope. This distribution is generally appropriate for making inferences about predictability, but it ignores useful information when the predictive variable's autocorrelation is close to one. Thus, it can substantially understate the significance of variables like DY , B/M , and E/P .

(b) The conditional tests in this paper are intuitive. If we knew ρ , the best estimate of β is the bias-adjusted estimator $\hat{\beta}_{\text{adj}} = \hat{\beta} - \gamma(\hat{\rho} - \rho)$. In practice, the true autocorrelation is unknown, but as long as DY is stationary then $\rho \approx 1$ is the most conservative assumption we can make because it yields the lowest estimate of β . The conditional tests are easy to apply: the necessary statistics can be estimated from OLS and, under the null, the test statistic has a student t distribution. When $\hat{\rho}$ is close to one, the conditional bias-adjusted slope will be higher than the unconditional estimate. Further, when $\hat{\beta}$ and $\hat{\rho}$ are highly correlated, the conditional variance will be much lower than the unconditional variance. Both of these effects help produce stronger tests of predictability.

(c) Empirically, incorporating the information in $\hat{\rho}$ can be quite important. I find strong evidence that DY predicts both equal- and value-weighted NYSE returns from 1946–2000. In the full sample and various subsamples, DY is typically significant at the 0.001 level, with many t -statistics greater than 3.0 or 4.0. The evidence for B/M and E/P ratios is somewhat weaker and, overall, they seem to

have limited forecasting power. Even when the statistics cannot reject the null, the conditional bias-adjusted slopes look much different than the unconditional estimates.

(d) The last few years of the sample have a large impact on the results. For the value-weighted index, adding 1995–2000 to the regressions reduces the OLS slope on DY by 59%, the slope on B/M by 61%, and the slope on E/P by 28%. However, the bias-adjusted estimates are less sensitive to the recent data. The estimates for equal-weighted returns actually increase with the addition of 1995–2000. This finding is explained by the sharp increase in the ratios' sample autocorrelations, which lowers the bias-adjustment needed in the conditional tests.

The conditional tests are based on a frequentist approach, but the methodology is similar, in some ways, to the Bayesian tests of [Stambaugh \(1999\)](#). Both approaches condition on the observed sample when deriving the distribution of β (in Bayesian tests) or $\hat{\beta}$ (in frequentist tests). The main advantage of the conditional test is that it assumes only that ρ is less than one; the Bayesian approach requires additional assumptions about investors' beliefs. The trade-off is that the conditional approach is useful only if the autocorrelation is close to one, while Bayesian tests are more general.

The evidence shows that information about ρ can be important. The only information used here is the stationarity of the predictive variable, but additional information about ρ could be incorporated into the tests. For example, suppose we have a sample beginning prior to the period in which we want to test for predictability. If DY 's autocorrelation is constant over the entire history, we could use the earlier data to help us infer whether the within-sample autocorrelation is above or below the true value. Also, the methodology could be generalized for tests with many predictive variables. The generalization should be straightforward when the stochastic nature of only one variable is a concern (for example, regressions with DY and T-bill rates). It is likely to be more difficult when all of the variables are correlated with returns.

Appendix

A.1. Estimating γ and σ_v

Section 2 shows that $\hat{\beta}_{\text{adj}} \equiv \hat{\beta} - \gamma(\hat{\rho} - \rho)$ has mean β and variance $\sigma_v^2(X'X)_{(2,2)}^{-1}$. The tests require estimates of γ and σ_v from the equation $\varepsilon_t = \gamma\mu_t + v_t$, where ε_t and μ_t are residuals in the predictive regression, $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$, and the AR1 model for DY , $x_t = \phi + \rho x_{t-1} + \mu_t$. To understand how sampling error in γ and σ_v affects the tests, it is useful to think about estimating β differently than described in the text. Re-write the predictive regression using $\varepsilon_t = \gamma\mu_t + v_t$:

$$r_t = \alpha + \beta x_{t-1} + \gamma\mu_t + v_t. \quad (\text{A.1})$$

Given an assumed ρ , this equation can be estimated because μ_t is observable. (More precisely, if we know ρ , then we observe $x_t - \rho x_{t-1} = \mu_t + \phi$; the value of ϕ does not

affect the estimate of β or γ and I ignore it in the remainder of the discussion.) The key observation, proved below, is that the estimate of β from (A.1) is identical to $\hat{\beta}_{\text{adj}}$ for a given ρ . Moreover, v_t is independent of x_t and μ_t at all leads and lags, so the OLS estimate of β from (A.1) has standard properties. Thus, when we estimate γ and σ_v , the tests should use the standard error from (A.1) and, under the null, the resulting t -statistic is truly a t -statistic, with $T - 3$ degrees of freedom.

From regression analysis, the estimate of β from (A.1), denoted $\hat{\beta}^M$, and the simple-regression estimate from the predictive regression are related by the formula:

$$\hat{\beta}^M = \hat{\beta} - \hat{\gamma}\lambda, \tag{A.2}$$

where λ is the slope coefficient in an auxiliary regression of μ_t on x_{t-1} :

$$\mu_t = c + \lambda x_{t-1} + \omega_t. \tag{A.3}$$

λ is the second element of $(X'X)^{-1}X'\mu$. Comparing this to Eq. (4b) in the text, we see that $\lambda = \hat{\rho} - \rho$. Substituting into (A.2) yields $\hat{\beta}^M = \hat{\beta} - \hat{\gamma}(\hat{\rho} - \rho)$. Note, also, that the estimate of γ from (A.1) is the same as the estimate from a regression of $\hat{\epsilon}_t$ on $\hat{\mu}_t$ (the sample residuals from the predictive regression and the AR1 model). This proves that $\hat{\beta}^M$ is identical to $\hat{\beta}_{\text{adj}}$.

The empirical tests rely on the argument that $\rho \approx 1$ is the most conservative choice for testing predictability (assuming DY is stationary). If γ is known, the proof is obvious because $\hat{\beta}_{\text{adj}}$ is minimized at $\rho = 1$ and the standard error of $\hat{\beta}_{\text{adj}}$ does not depend on ρ .⁹ When γ is unknown, the proof is harder because the standard error becomes a function of ρ . To see this, re-write (A.1) to emphasize that μ_t is a function of ρ :

$$r_t = \alpha + \beta x_{t-1} + \gamma(x_t - \rho x_{t-1}) + v_t. \tag{A.4}$$

As we change ρ , we simply add or subtract x_{t-1} from the second regressor. Doing so affects only the estimate of β , not the residuals v_t or the estimate of γ . The standard error of $\hat{\beta}_{\text{adj}}$ in this regression equals $\hat{\sigma}_v/(T^{1/2}\sigma_s)$, where T is time-series length and σ_s^2 is the residual variance when x_{t-1} is regressed on the other explanatory variables: $x_{t-1} = c + \delta\mu_t + s_t$. The variance of s_t is

$$\sigma_s^2 = \sigma_x^2 - \delta^2\sigma_\mu^2 = \sigma_x^2 - \frac{[(\hat{\rho} - \rho)\sigma_x^2]^2}{\sigma_{\mu,\text{OLS}}^2 + (\hat{\rho} - \rho)^2\sigma_x^2}. \tag{A.5}$$

In this equation, all of the variances are sample statistics, not adjusted for degrees of freedom. The second equality uses the fact that $\mu_t = \hat{\mu}_t + (\hat{\rho} - \rho)x_{t-1}$, where $\hat{\rho}$ and $\hat{\mu}_t$ are OLS estimates, and $\hat{\mu}_t$ and x_{t-1} are uncorrelated by construction. Simplifying yields

$$\sigma_s^2 = \sigma_x^2 \left[\frac{\sigma_{\mu,\text{OLS}}^2}{\sigma_{\mu,\text{OLS}}^2 + (\hat{\rho} - \rho)^2\sigma_x^2} \right]. \tag{A.6}$$

⁹In this case, the test statistic has $T - 2$ degrees of freedom. The standard error of β_{adj} is the square root of $\hat{\sigma}_v^2(X'X)^{-1}_{(2,2)}$, where $\hat{\sigma}_v^2$ is the sample variance of v_t . In practice, γ can be estimated very precisely and estimation error in $\hat{\gamma}$ has almost no effect on the tests. For example, in full-sample regressions with VWNY, $\hat{\gamma}$ is -90.4 with a standard error of 1.1.

This equation shows that σ_x^2 , and consequently the standard error of $\hat{\beta}_{\text{adj}}$, is a function of $(\hat{\rho} - \rho)^2$. This effect is relatively small for reasonable values of ρ . Formally, let $k = T^{1/2}\sigma_x\sigma_{\mu,\text{OLS}}/\hat{\sigma}_v$. The t -statistic from (A.1) is

$$t_{\beta=0} = k \frac{\hat{\beta} - \hat{\gamma}(\hat{\rho} - \rho)}{[\sigma_{\mu,\text{OLS}}^2 + (\hat{\rho} - \rho)^2\sigma_x^2]^{1/2}}. \quad (\text{A.7})$$

The numerator is decreasing in ρ , while the denominator is increasing in $(\hat{\rho} - \rho)^2$. This function has a unique local extremum at $\rho = \hat{\rho} + (\hat{\gamma}\sigma_{\mu,\text{OLS}}^2/\hat{\beta}\sigma_x^2)$. Assuming $\hat{\beta} > 0$ and $\hat{\gamma} < 0$, this is a global maximum and occurs at some $\rho < \hat{\rho}$. Notice, also, that the t -statistic converges to $-k\hat{\gamma}/\sigma_x$ from above as $\rho \rightarrow -\infty$. Together, these facts imply that the t -statistic is minimized at $\rho = 1$ as long as the asymptotic value, $-k\hat{\gamma}/\sigma_x$, is greater than the value at $\rho = 1$. Substituting for k , the asymptotic value is $-\hat{\gamma}T^{1/2}\sigma_{\mu,\text{OLS}}/\hat{\sigma}_v$, which happens to be the t -statistic for testing whether $-\hat{\gamma}$ equals zero (i.e., whether shocks to returns and shocks to DY are related). In applications, this t -statistic will be very large and much greater than the value of $t_{\beta=0}$ at $\rho = 1$ (e.g., it is 82.1 in the full-sample VWNY regressions).

A.2. Power

Section 2.4 argues that we cannot say, *ex ante*, whether the conditional test (based on the assumption that $\rho \approx 1$) or the unconditional test (based on the marginal distribution of $\hat{\beta}$) is better. The conditional test has greater power if ρ is close to one, but the opposite is true once ρ drops below some level that depends on the other parameters. If both tests are used, the joint significance level can be based on the modified Bonferroni p -value suggested in Section 2.4. This section explores the power of the tests, and the modified Bonferroni p -value, in greater detail.

Some basic facts are immediate, though formalized below. First, it is well known that the unconditional test has relatively low power: stock returns are volatile, and the standard error in predictive regressions is high. The asymmetry in the marginal distribution of $\hat{\beta}$ (the long right tail in Panel A of Fig. 1) accentuates the problem. Second, the conditional test imposes a trade-off: conditioning on $\hat{\rho}$ dramatically lowers the standard error of $\hat{\beta}$, which increases power, but the assumption that $\rho \approx 1$ is very conservative, biasing downward the slope estimate if ρ is really less than one. In particular, if we assume $\rho = 1$, then $\hat{\beta}_{\text{adj}} = \hat{\beta} - \gamma(\hat{\rho} - 1)$ has mean $\beta - \gamma(\rho - 1)$. The estimate is unbiased only if $\rho = 1$ and otherwise biased downward ($\gamma < 0$), understating predictability and reducing the power of the tests. It should be clear that the conditional test is more powerful if ρ is close to one, but power drops to zero as ρ becomes smaller.

Table A.1 explores these ideas. It shows rejection rates for Stambaugh's (1999) unconditional test, the conditional test that assumes $\rho \approx 1$, and a joint test based on the modified Bonferroni p -value introduced in Section 2.4. The rejection rates come from 5,000 Monte Carlo simulations (not really necessary for the ' $\rho \approx 1$ ' results). The simulations are calibrated to the VWNY regressions on log DY for 1946–1972 (see Table 3), except that β and ρ are set to the values indicated in the table. β varies from

Table A.1

Power and significance levels

The table reports rejection rates for three tests of predictability: (i) Stambaugh's (1999) unconditional test, based on the marginal distribution of the OLS slope; (ii) the conditional test, which assumes that $\rho \approx 1$, and (iii) an overall, or joint, test based on the modified Bonferroni p -value, as outlined in Section 2.4. The model for returns is $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$, where x_t follows an ARI process, $x_t = \phi + \rho x_{t-1} + \mu_t$. Given an assumed β and ρ , the table shows the probability that each test rejects that $\beta = 0$ at a nominal 5% significance level. The probabilities are estimated from 5,000 Monte Carlo simulations. The true parameters in the simulations are the OLS estimates for value-weighted NYSE returns regressed on $\log DY$ from 1946–1972 (see Table 3), except that β and ρ are set as indicated in the table.

Test	β	ρ												
		0.999	0.997	0.995	0.993	0.991	0.989	0.987	0.985	0.983	0.981	0.979	0.977	0.975
Stambaugh	0.0	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050	0.050
	0.4	0.071	0.070	0.067	0.070	0.072	0.073	0.073	0.073	0.071	0.071	0.071	0.071	0.071
	0.8	0.104	0.097	0.096	0.099	0.097	0.098	0.100	0.103	0.101	0.101	0.100	0.101	0.101
	1.2	0.151	0.137	0.137	0.134	0.137	0.137	0.137	0.143	0.136	0.138	0.138	0.139	0.136
	1.6	0.209	0.193	0.190	0.187	0.186	0.188	0.189	0.193	0.186	0.186	0.183	0.183	0.181
$\rho \approx 1$	0.0	0.027	0.009	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.301	0.117	0.040	0.014	0.004	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.8	0.712	0.516	0.301	0.141	0.060	0.022	0.008	0.003	0.001	0.000	0.000	0.000	0.000
	1.2	0.914	0.826	0.678	0.500	0.309	0.164	0.077	0.032	0.013	0.004	0.002	0.001	0.000
	1.6	0.976	0.952	0.901	0.808	0.667	0.496	0.326	0.182	0.091	0.044	0.019	0.007	0.003
Joint test	0.0	0.041	0.032	0.034	0.035	0.035	0.036	0.037	0.038	0.040	0.041	0.043	0.043	0.044
	0.4	0.242	0.106	0.058	0.047	0.047	0.049	0.050	0.052	0.054	0.055	0.056	0.060	0.062
	0.8	0.659	0.445	0.250	0.137	0.085	0.069	0.062	0.065	0.067	0.071	0.074	0.077	0.079
	1.2	0.897	0.784	0.627	0.448	0.279	0.170	0.114	0.092	0.090	0.092	0.097	0.102	0.107
	1.6	0.975	0.944	0.880	0.769	0.627	0.462	0.304	0.201	0.151	0.129	0.124	0.128	0.132

0.0 to 1.6 in increments of 0.4, and ρ varies from 0.999 to 0.975 in increments of -0.002 . Given historical data, the ranges span reasonable values for the parameters. The OLS estimate in Table 3 is 1.42 for β and 0.993 for ρ .

The table shows several key results. First, the conditional test by itself and the joint test based on the modified Bonferroni p -value are conservative in the sense that they reject too little under the null (the rows with $\beta = 0.0$). The rejection rates are always less than the nominal 5% significance level; the conditional p -value is accurate if ρ is very close to 1, while the joint p -value is accurate if ρ is either very close to or very far from one.

Second, the power of the conditional test falls quickly as ρ gets smaller and grows quickly as β gets larger. Power is high if ρ is close to one or β is large. For example, if β is 1.2, the rejection rates are between 31% and 91% if $\rho \geq 0.991$, and then drop rapidly for lower values of ρ . If $\rho < 0.98$, the rejection rates are close to zero for all values of β considered. In contrast, the power of the unconditional test is modest for all values of β but has the key advantage that it is stable across values of ρ . The rejection rates are around 10% for $\beta = 0.8$, 14% for $\beta = 1.2$, and 19% for $\beta = 1.6$.

Finally, the joint test—which relies on both the conditional and unconditional tests—seems to do a good job combining the advantages of the two individual tests.

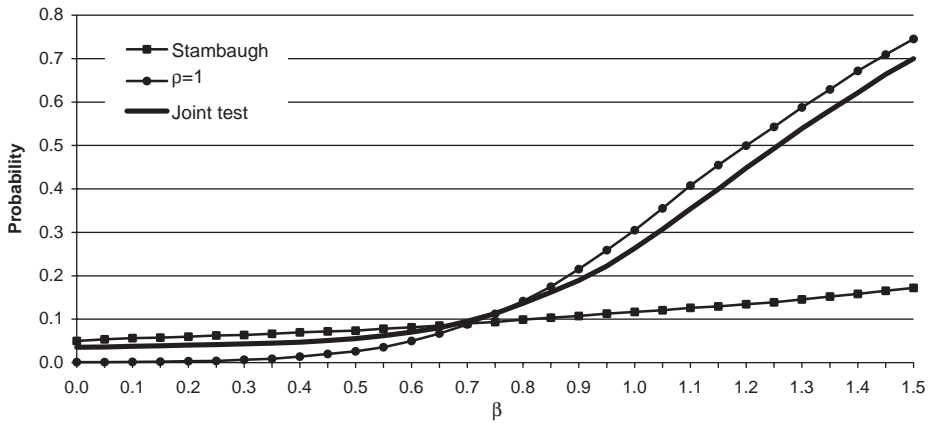


Fig. A.1. Tests of predictability: statistical power. The figure plots rejection rates for three tests of predictability: (i) Stambaugh's (1999) unconditional test, based on the marginal distribution of the OLS slope; (ii) the conditional test which assumes that $\rho \approx 1$, and (iii) an overall, or joint, test based on the modified Bonferroni p -value, as outlined in Section 2.4. The model for returns is $r_t = \alpha + \beta x_{t-1} + \varepsilon_t$, where x_t follows an AR1 process, $x_t = \phi + \rho x_{t-1} + \mu_t$. The figure shows the probability that each test rejects that $\beta = 0$ at a nominal 5% significance level. The probabilities are estimated from 5,000 Monte Carlo simulations. The true parameters in the simulations are the OLS estimates for value-weighted NYSE returns regressed on $\log DY$ from 1946–1972, except that β varies from 0.0 to 1.6 (the autocorrelation equals 0.993).

It has fairly high power when either β or ρ is high, and the rates do not drop off as quickly as they do for the conditional test when β or ρ becomes smaller. The trade-off is that it sacrifices some power, relative to the unconditional test, at smaller values of β and ρ .

Fig. A.1 illustrates these results, focusing on how power changes across values of β . It shows rejection rates for a given value of ρ (0.993, equal to the OLS estimate) as β increases from 0.00 to 1.60. If β is small, the conditional test has essentially no power because the assumption that $\rho \approx 1$ is too conservative. However, as β gets large, the power of the conditional test rises dramatically. It reaches 81% when $\beta = 1.6$, while the power of the unconditional test rises to only 19%. Perhaps more important, the figure shows that using both together—that is, the joint test—combines the advantages of each at the cost of sacrificing some power to the unconditional test at low values of β and some power to the conditional tests at high values of β .

References

- Ang, A., Bekaert, G., 2002. Stock return predictability: is it there? Working Paper, Columbia University, New York.
- Blanchard, O., Watson, M., 1982. Bubbles, rational expectations, and financial markets. In: Watchel, P. (Ed.), *Crises in the Economic and Financial Structure*. Lexington Books, Lexington, MA, pp. 295–315.
- Campbell, J., 1987. Stock returns and the term structure. *Journal of Financial Economics* 18, 373–399.

- Campbell, J., Shiller, R., 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195–228.
- Campbell, J., Yogo, M., 2003. Efficient tests of stock return predictability. Working Paper, Harvard University, Cambridge, MA.
- Diba, B., Grossman, H., 1988. Explosive rational bubbles in stock prices? *American Economic Review* 78, 520–530.
- Fama, E., French, K., 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3–25.
- Fama, E., French, K., 2002. The equity premium. *Journal of Finance* 57, 637–659.
- Fama, E., Schwert, G.W., 1977. Asset returns and inflation. *Journal of Financial Economics* 5, 115–146.
- Flood, R., Hodrick, R., 1986. Asset price volatility, bubbles, and process switching. *Journal of Finance* 41, 831–842.
- Flood, R., Hodrick, R., 1990. On testing for speculative bubbles. *Journal of Economic Perspectives* 4, 85–101.
- Goetzmann, W., Jorion, P., 1993. Testing the predictive power of dividend yields. *Journal of Finance* 48, 663–679.
- Hamilton, J., Whiteman, C., 1985. The observable implications of self-fulfilling expectations. *Journal of Monetary Economics* 16, 353–373.
- Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357–386.
- Kendall, M., 1953. The analysis of economic time series, part I: prices. *Journal of the Royal Statistical Society* 96, 11–25.
- Kothari, S.P., Shanken, J., 1997. Book-to-market, dividend yield, and expected market returns: a time-series analysis. *Journal of Financial Economics* 44, 169–203.
- Kothari, S.P., Shanken, J., Sloan, R., 1995. Another look at the cross-section of expected stock returns. *Journal of Finance* 50, 185–224.
- Lamont, O., 1998. Earnings and expected returns. *Journal of Finance* 53, 1563–1587.
- Loewenstein, M., Willard, G., 2000. Rational equilibrium asset-pricing bubbles in continuous trading models. *Journal of Economic Theory* 91, 17–58.
- Mankiw, N.G., Shapiro, M., 1986. Do we reject too often? small sample properties of tests of rational expectations models. *Economic Letters* 20, 139–145.
- Nelson, C., Kim, M., 1993. Predictable stock returns: the role of small sample bias. *Journal of Finance* 48, 641–661.
- Pontiff, J., Schall, L., 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics* 49, 141–160.
- Shiller, R., 1984. Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 457–498.
- Stambaugh, R., 1986. Bias in regressions with lagged stochastic regressors. Unpublished Manuscript, University of Chicago, Chicago, IL.
- Stambaugh, R., 1999. Predictive regressions. *Journal of Financial Economics* 54, 375–421.
- Tirole, J., 1982. On the possibility of speculation under rational expectations. *Econometrica* 50, 1163–1181.
- Tirole, J., 1985. Asset bubbles and overlapping generations. *Econometrica* 53, 1071–1100.
- West, K., 1988. Bubbles, fads, and stock price volatility tests: a partial evaluation. *Journal of Finance* 43, 639–656.