MIT Sloan School of Management

Working Paper 4374-02
August 2002

PREDICTING RETURNS WITH FINANCIAL RATIOS

Jonathan Lewellen

Predicting returns with financial ratios

Jonathan Lewellen

MIT Sloan School of Management
50 Memorial Drive E52-436, Cambridge, MA 02142
(617) 258-8408
lewellen@mit.edu

First draft: March 2000
Current version: August 2002

I am grateful to John Campbell, Ken French, Sydney Ludvigson, Bill Schwert, Jay Shanken, Rob Stambaugh, Tuomo Vuolteenaho, and two anonymous referees for helpful comments and suggestions. The paper has benefited from workshops at Chicago, Dartmouth, Duke, MIT, Michigan, NBER, and Grantham, Mayo, Van Otterloo, and Co. Financial support from the MIT / Merrill Lynch Partnership is gratefully acknowledged.

Predicting returns with financial ratios

Abstract

This article provides a new test of the predictive ability of aggregate financial ratios. Predictive regressions are subject to small-sample biases, but the correction in previous studies can substantially understate forecasting power. Dividend yield predicts aggregate market returns from 1946 – 2000, as well as in various subperiods. Book-to-market and the earnings-price ratio predict returns during the shorter 1963 – 2000 sample. The evidence remains strong despite the unusual price run-up in recent years.

Predicting returns with financial ratios

1. Introduction

Nearly fifty years ago, Kendall (1953) observed that stock prices seem to wander randomly over time. Kendall, and 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, book-to-market, and the earnings-price ratio (see, e.g., Fama 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, each ratio measures price relative to 'fundamentals.' Because price is high when expected returns are low, and vice versa, 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 important effects on empirical tests.

This paper provides new tests of the forecasting ability of aggregate DY, B/M, and E/P. I focus primarily on DY because it has received the most attention in the literature. I also focus exclusively on short-horizon tests – monthly returns regressed on lagged DY – to avoid the complications arising from overlapping returns. Previous studies provide weak evidence that DY forecasts 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, 1999) 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 an AR1 process. He reports a one-sided p-value of 0.15 when NYSE returns are regressed on lagged DY from 1952 – 1996.[1]

In this paper, I show that Stambaugh's (1999) small-sample distribution, as well as simulations common in the literature, can substantially understate the forecasting ability of the financial ratios. Although the small-sample adjustment is generally appropriate, we can sometimes improve upon it by incorporating information about DY's sample autocorrelation. The sample autocorrelation is strongly correlated with the slope estimate in the predictive regression, so any information conveyed by the autocorrelation helps produce a more powerful tests of predictability. Incorporating this information into empirical tests has two effects: (1) the slope coefficient is often larger than Stambaugh's estimate; and (2) the standard error of the estimate is much lower. In combination, the two effects can substantially raise the power of empirical tests.

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

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

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

where r_t is the stock return and x_{t-1} is the dividend yield (or other financial ratio). Equation (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:

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

where η is a random error with mean zero and γ is a negative constant. Empirical tests are typically based

[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 (2001).

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 throws out any information we have about $\hat{\rho}-\rho$. Specifically, if we are willing to assume that DY is stationary, so that ρ is at most one, 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 is less than the standard bias-adjustment if $\hat{\rho}$ is close to one. When this occurs, 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 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 emphasize that these estimates are conservative, calculated under the assumption that $\rho \approx 1$; the estimates are biased downward if ρ is actually less than one.) Predictability is also strong in subperiods. 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 affect the empirical results. DY reached a new low for the sample in May 1995, predicting that returns going forward should be far below average. In fact, 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 not sensitive 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, finding stronger evidence of predictability after observing the recent data.

B/M and E/P also have significant predictive power, though the evidence is less reliable. The tests with these variables begin in 1963 when Compustat becomes available. From 1963 – 1994, B/M and E/P forecast both equal- and value-weighted NYSE indices, with p-values less than 0.050. When more recent data is included, however, the ratios seem to predict only the equal-weighted index. The evidence is much stronger than previous studies. Kothari and Shanken (1997) and Pontiff and Schall (1998) find that B/M has little predictive power during this period, and Lamont (1999) finds no evidence that E/P (by itself) predicts quarterly returns from 1947 – 1994.

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

2. Predictive regressions

Predictive regressions are ubiquitous 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

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

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 the 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:

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

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 x_t is correlated with ε_t in the predictive regression, violating one of the assumptions of OLS (which requires independence at all leads and lags). For simplicity, I assume the variables are normally distributed.

Before continuing, I should briefly discuss the stationarity assumption. The empirical tests depend on the assumption that ρ cannot be 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. x_t will be stationary if log dividends and log prices are cointegrated, implying that, in the long run, dividends and prices must 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]

It is useful to note that DY might exhibit other forms of nonstationarity. 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. The regression in eq. (3b) can then be thought of as an restricted version of a model with time-varying parameters (i.e., $x_t = \phi_t + \rho\, x_{t-1} + \mu_t$). The empirical tests should be relatively insensitive to this type of nonstationarity as long as ρ remains below one.

---

[2] Blanchard and Watson (1982), Tirole (1982, 1985), and Loewenstein and Willard (1999) discuss the theory of rational bubbles and Hamilton and Whiteman (1985), Flood and Hodrick (1986, 1990), Diba and Grossman (1988), and West (1988) provide empirical evidence.

2.2. Properties of OLS

Denote the matrix of regressors as X, the coefficient vectors as b = (α, β) and p = (ϕ, ρ), and the residual vectors as ε and μ. The OLS estimates of eqs. (3a) and (3b) are then

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

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

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, note that we can write $\varepsilon_t = \gamma\ \mu_t + \nu_t$, where γ = cov(ε, μ) / var(μ). Substituting into eq. (4a) yields

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

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

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

$$E[\hat{\beta} - \beta] = \gamma\ E[\hat{\rho} - \rho] . \tag{6}$$

Sample autocorrelations are biased downward by approximately $-(1+3\rho)/T$, 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 (1999) discusses these results in detail.

The empirical tests below will also use the conditional distribution of $\hat{\beta}$ given $\hat{\rho}$. Equation (5) shows that, although $\hat{\beta}$ and $\hat{\rho}$ are not bivariate normal due to the irregularities in $\hat{\rho}$, $\hat{\beta}$ is normally distributed conditional on $\hat{\rho}$.[3] This observation follows from the definition of η in eq. (5). The conditional expectation of $\hat{\beta}$ is

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

which I refer to, informally, as the 'realized bias' in $\hat{\beta}$. (The conditional variance of $\hat{\beta}$ is given by var(η), defined above.) Equation (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 are based on the conditional distribution of $\hat{\beta}$. The idea is simple. Even though we do not know $\hat{\rho} - \rho$, which is required for the conditional distribution, 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 $\hat{\beta}$. To be precise, define the bias-adjusted estimator

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

Given the true ρ, this estimator is normally distributed with mean β and variance $\sigma_v^2 (X'X)^{-1}_{(2,2)}$. We do not know ρ, of course, but setting $\rho \approx 1$ is the most conservative assumption we can make for testing predictability: the bias in eq. (7) is maximized, and the estimator in eq. (8) is minimized, if we assume $\rho \approx 1$. If $\hat{\beta}_{adj}$ is significantly different from zero given this assumption, then it must be even more significant given the true value of ρ.[4]

There is a second way to think about this test. The analysis above shows that the sampling errors in

[3] Technically, I am conditioning on both $\hat{\rho}$ and X. The matrix X is important because of its impact on the variance of η (see eq. 5). If we conditioning only on $\hat{\rho}$, the conditional distribution of β is a mixture of normal distributions, all with mean zero but different variances. Conditioning on X is similar to standard OLS with stochastic regressors.

[4] To implement the tests, we also need to estimate γ and σ_v from $\varepsilon_t = \gamma\,\mu_t + \nu_t$. The estimators are the OLS estimates using the sample values of ε_t and μ_t. The Appendix describes how estimation error in γ and σ_v affects the statistical tests.

$\hat{\beta}$ 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 $\rho < 1$.

Previous studies have focused on the unconditional distribution of $\hat{\beta}$. They implicitly assume we have no information about $\hat{\rho} - \rho$. That assumption is fine when $\hat{\rho}$ is low 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 $T = 300$. The unconditional bias in $\hat{\rho}$ is approximately $E[\hat{\rho} - \rho] = -0.016$. However, given the observed autocorrelation, the *minimum* possible value of $\hat{\rho} - \rho$ is actually -0.010. The conditional tests use this information to improve inferences about β. More generally, the conditional approach 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 might be useful if the autocorrelation is a bit lower, but it depends on the underlying parameters.)

Figure 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 the $\hat{\beta}$ 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.

An unfortunate feature of the tests is that we cannot say, ex ante, whether the conditional or unconditional approach 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 (depending on the parameters). Ideally, we could choose between the tests, in advance, based on either economic theory or pre-sample evidence. In practice, the decision will likely depend on the sample autocorrelation, and the significance level should recognize that we 'search' for the stronger test after looking at the data:

(i) If, ex ante, the researcher believes that ρ is far from one and uses only the unconditional test, then no adjustment is necessary. The p-value from the unconditional tests is appropriate.

(ii) If, ex ante, the researcher believes that ρ might be close to one and relies on both tests, then an overall p-value (probability of rejecting using either test) is given by twice the smaller p-value. The result is Bonferroni's upper bound on the true significance level.[5]

The p-value in (ii) assumes that we apply both tests regardless of ρ. The bound is tight if ρ is close to one, but otherwise quite conservative: the conditional test will rarely reject if ρ is low, so doubling the smaller (unconditional) p-value is not really necessary. A better overall p-value would recognize this fact. It is difficult to derive an alternative p-value, but simulations suggest the following gives a tighter upper bound:

(ii′) If a researcher uses both tests, then an upper bound on the overall p-value 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 $\rho = 1$ (based on the sampling distribution of $\hat{\rho}$).[6]

The first part, 2P, is just the Bonferroni upper bound; the second part, P + D, recognizes that doubling P is

---

[5] Suppose 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 or B) = Pr(A) + Pr(B) – Pr(A and B) < 2P.

[6] The bound is derived from asking: what is the probability either test rejects with probability P and the sample autocorrelation has p-value D under the null? Let A, B, and C be the rejection regions for the conditional test, unconditional test, and unit root test, respectively. We have: Pr((A or B) and C) = Pr((A and C) or (B and C)) < Pr(A) + min(Pr(B), Pr(C)) = P + min(P, Pr(C)). If $\rho = 1$, the bound obtains because 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. I cannot prove that the bound holds for all ρ, but simulations suggest that it does. I ran simulations calibrated to the DY regressions, allowing ρ to vary from 0.9 to 0.9999. I calculated min(2P, P+D) for each simulation and assumed we reject if the value is below 0.05. Under the null, the true rejection rate was less than or equal to 5% for all ρ.

too conservative if we reject $\rho = 1$ (that is, if D is small). This alternative bound is important because it blurs the distinction between (i) and (ii). In particular, (ii′) says 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, even if we also considered the conditional test. The adjustment for searching becomes larger if $\hat{\rho}$ is close to one, with a maximum provided by Bonferroni.

The conditional approach is a natural addition to Stambaugh's (1999) unconditional test. It is important to emphasize that his analysis gives the better estimate of β unless the autocorrelation is close to one. Also, Stambaugh presents Bayesian tests in which he sometimes assumes that $\rho < 1$. He shows that this can lead to stronger rejections of the null. My contribution is to show that the constraint can also improve inferences in a frequentist setting. In fact, the approach here is similar in many ways to Stambaugh's Bayesian analysis, in that both condition on observed data in deriving the distribution of β (or $\hat{\beta}$ here). The tests are essentially 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 even stronger rejections of the null.

3. Data and descriptive statistics

I use the methodology outlined above to test whether DY, B/M, and E/P forecast stock returns. Prices and dividends come from the Center for Research in Security Prices (CRSP) database. Earnings and book equity 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 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. The 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 extend from January 1946 – December 2000. I omit the Depression era because the properties of returns 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 1963 – 2000 when Compustat data is available. B/M is the ratio of book equity in the previous fiscal year to market equity in the previous month. E/P is the ratio of operating earnings (before depreciation) to market value. 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 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, Shanken, and Sloan, 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

[7] Log 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 calculated from net income (standard deviation of 0.062 compared with 0.049 for operating earnings). In addition, the residuals are less highly correlated with returns, –0.66 vs. –0.86. Net income seems to vary independent of price movements more than operating earnings, which might indicate additional noise in the process. In any case, the predictive power of the two series is similar and, for simplicity, I report only the tests using operating earnings (the results for net income are marginally weaker).

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 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

I estimate predictive regressions using DY, B/M, and E/P. The tests initially focus on DY because it has been studied most in the literature. The last few years of the sample receive special attention because they have a large impact on the results.

4.1. Predicting with dividend yield

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

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

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

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 Tbill 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 reference point. The row labeled 'Stambaugh' shows estimates based on Stambaugh's (1999) small-sample distribution. The point estimate is bias-adjusted using the unconditional bias and the p-value is based on the unconditional distribution, both 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.

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

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

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 tests understate the predictive power of DY. The variance of $\hat{\beta}_{adj}$ is $\sigma_v^2 (X'X)^{-1}$. To implement the tests, we need to estimate γ and σ_v^2 from $\varepsilon_t = \gamma\ \mu_t + \nu_t$, where ε and μ are the residuals in eq. (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.[8]

Table 2 provides strong evidence of predictability. Consider, first, the nominal return on the value-weighted index. 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 a one-sided p-value of 0.308. The conditional test, in the third row, gives a much different picture. Assuming that $\rho \approx 1$, the 'realized bias' in the predictive regression is at most 0.25, approximately 65% lower than the unconditional bias. The bias-adjusted slope is 0.66 with a t-statistic of 4.67, significant at the 0.000 level.

[8] The tables report raw p-values for each test. As discussed earlier, an overall test of predictability should recognize that we perform multiple tests (see Section 2.3 for details). Doing so would not affect any of the inferences in this paper.

Notice that the conditional standard error is much lower than the standard error estimated from Stambaugh's distribution, 0.14 vs. 0.67.[9] The strong significance in the conditional test is largely attributable to this difference.

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 ignore the information contained in the sample autocorrelation of DY, which lowers the point estimate and raises the standard error when $\hat{\rho}$ is close to one. Surprisingly, the conditional tests in Table 2 find that DY is more significant than suggested by OLS. The bias-adjusted slope is lower than the OLS estimate, but the information conveyed by $\hat{\rho}$ has a more important effect on the standard error. As a result, the stochastic properties of DY actually strengthen the case for predictability.

The regressions for equal-weighted NYSE returns confirm these findings. The bias-adjusted slope for EWNY, 1.12, is larger than the estimate for VWNY and suggests significant time-variation in expected returns (p-value of 0.000). The table shows that nominal and excess returns produce similar estimates during this period. Also, in all regressions, the explanatory power of DY is low as measured by the adjusted R^2. Thus, while DY tracks significant variation in expected returns, it only explains a small fraction of monthly stock volatility.

I should point out 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. The regressions with nominal returns are similar to those in Table 2, while the estimates for excess returns drop considerably, in part because the correlation between DY and Tbill rates is

[9] The standard error is relevant only for testing a lower bound on the predictive slope (i.e., for testing the null of no predictability). The standard error on the upside is higher because we have only a one-sided bound on the DY's autocorrelation. Thus, β might be substantially larger than suggested by a symmetric confidence interval around the conditional estimate.

sensitive to the starting date of the sample.[10]

Table 3 reports results for the first and second halves of the sample, 1946 – 1972 and 1973 – 2000. Even with a fairly short sample, the tests strongly reject the null in most cases. For 1946 – 1972, DY predicts excess returns on both indices. Focusing on the value-weighted index, the bias-adjusted estimates for nominal and excess returns, 0.84 and 1.16 (p-values of 0.001 and 0.000), are similar to those in the full sample. In the second half of the sample, the estimate for raw returns drops slightly, to 0.64, while the slope for excess VWNY falls by nearly 75% (but remains statistically significant). The coefficients are especially large for the equal-weighted index from 1973 – 2000. The bias-adjusted slope for excess EWNY, 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 VWNY 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 is $\hat{\rho}$ 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. If ρ is truly less than one, the conditional estimate of β and its t-statistic are both biased downward. Although it is difficult to justify any other value for ρ, it might be useful to explore alternative values. Consider the results for nominal EWNY from 1946 – 1972. The p-value drops from 0.147 if we assume that $\rho \approx 1$ to 0.004 if we assume

[10] For example, the correlation is 0.00 for 1946 – 2000 but 0.24 for 1952 – 2000. Regressions using excess returns together with raw DY 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 Tbill rates (details available on request).

that $\rho = 0.993$ (the sample autocorrelation). The p-value would be less than 0.050 for any $\rho < 0.997$. While I am not suggesting that DY is truly significant in this regression, the reported p-values do appear to be quite conservative.

Bayesian methods provide a more rigorous way to incorporate uncertainty about ρ.[11] In particular, suppose an investor begins with no information about β, and consider three different beliefs about ρ. If the investor believes $\rho \approx 1$ with certainty, the conditional p-values in Table 3 equal the posterior probability that $\beta \leq 0$. For nominal EWNY, this probability is 0.147. If, instead, the investor begins with a flat prior for $\rho \leq 1$, the posterior probability drops 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. Returns during these years moved strongly opposite the predictions of the model. DY was extremely low, dropping from 2.9% in January 1995 to 1.5% in December 2000, while the value-weighted index grew 195%. Because this period is so unusual, I briefly consider its effect on the results.

Table 4 shows results for 1946 – 1994 together with corresponding estimates for the full sample (the same as Table 2). I only report 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 only 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

[11] The Bayesian posterior probability can be found by integrating the conditional p-value (conditional on a given autocorrelation) over the posterior distribution of ρ. The analysis here is like that of Stambaugh (1999).

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 from 1995 – 2000.

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 it to movement 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 – 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 is 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

The evidence above shows that DY predicts returns much more strongly than other studies have found. B/M and E/P have also been used with limited success to predict returns. Tables 5 and 6 explore the predictive power of these variables. I report regressions both for the full sample, 1963 – 2000, and the truncated sample ending in 1994.

B/M has some power to forecast returns, 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 p-values of 0.022 and 0.017. When returns for 1995 – 2000 are added, B/M predicts both raw and excess returns on the equal-weighted index, but has little predictive power for VWNY. The estimates for the equal-weighted index 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 results for DY, Table 5 shows that conditional tests provide much stronger evidence of predictability than unconditional tests. The differences are dramatic for nominal returns in the full sample. Incorporating the information in $\hat{\rho}$, the p-value for VWNY drops from 0.515 to 0.142 and the p-value for EWNY drops from 0.266 to 0.005. For both indices, the conditional bias is less than half the unconditional bias. The difference between the two tests is also revealed by comparing the full and truncated samples. The regressions for excess EWNY are especially interesting. The OLS slope estimate is 1.79 for the sample ending in 1994 and 1.14 for the sample ending in 2000. However, the conditional bias-adjusted slope actually *increases* slightly, from 0.57 to 0.68. That result mirrors the evidence for DY in Table 4.

Table 6 replicates the tests using E/P. The results 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 strengthens the case for predictability.

5. Summary

The literature on stock return predictability has evolved considerably over the last twenty years. Empirical tests initially produced strong evidence that market returns are predictable, especially over long horizons. Later research questioned these findings, suggesting 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. This paper provides new tests of their predictive ability, emphasizing four main points:

(a) Stambaugh (1999) and Mankiw and Shapiro (1986) consider the 'unconditional' distribution of $\hat{\beta}$. 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. 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 $\hat{\beta}_{adj} = \hat{\beta} - \gamma(\hat{\rho} - \rho)$. This estimator is normally distributed with mean β and variance $\sigma_v^2(X'X)^{-1}$. Although we do not know true autocorrelation, $\rho \approx 1$ is the most conservative assumption we can make because it yields the lowest estimate of β. The conditional tests are also 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 $\hat{\rho}$ can be quite important. I find strong evidence that DY predicts returns. The tests examine NYSE equal- and value-weighted indices from 1946 – 2000. In both 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 weaker than for DY, but stronger than previous studies. Overall, B/M and E/P appear to have limited forecasting ability. 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 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 can 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 Tbill rates). It is likely to be more difficult when all of the variables are correlated with returns.

Appendix

Section 2 shows that $\hat{\beta}_{adj} = \hat{\beta} - \gamma(\hat{\rho} - \rho)$ is normally distributed with mean β and variance $\sigma_\nu^2 (X'X)^{-1}_{(2,2)}$. To implement the tests, we need to estimate γ and σ_ν^2 in $\varepsilon_t = \gamma\ \mu_t + \nu_t$. The estimates are the OLS estimates using the sample values of ε and μ from the predictive regression, $r_t = \alpha + \beta\ x_{t-1} + \varepsilon_t$, and AR1 model for DY, $x_t = \phi + \rho\ x_{t-1} + \mu_t$. This appendix describes how sampling error in γ and σ_ν^2 affects the tests.

To analyze the problem, it is useful to think about estimating β somewhat differently than described in the text. Re-write the predictive regression using $\varepsilon_t = \gamma\ \mu_t + \nu_t$:

$$r_t = \alpha + \beta\ x_{t-1} + \gamma\ \mu_t + \nu_t. \tag{A.1}$$

Given ρ, this equation can be estimated since $\mu_t = x_t - \rho\ x_{t-1}$ is observable.[12] Moreover, ν_t is independent of x_t and μ_t at all leads and lags. It follows that (A.1) satisfies OLS, so the estimate of β has all the usual properties. I now show that the estimate of β from (A.1) is identical to $\hat{\beta}_{adj}$, thus providing the sampling distribution of $\hat{\beta}_{adj}$.

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 paper, we see that $\lambda = \hat{\rho} - \rho$. Substituting into (A.2) yields $\hat{\beta}^M = \hat{\beta} - \hat{\gamma}\,(\hat{\rho} - \rho)$. Also, the estimate of γ from (A.1) is the same as the estimate from a regression of $\hat{\varepsilon}_t$ on $\hat{\mu}_t$ (the sample residuals from the predictive regression and AR1

---

[12] 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 (it affects only the intercept), and I ignore it in the remainder of the discussion.

model). Therefore, $\hat{\beta}^M$ is identical to $\hat{\beta}_{adj}$. These results imply that, when γ and σ_v are unknown, inference should be based on the t-statistic from (A.1). Under the null, it has a t-distribution with T – 3 degrees of freedom.

Using (A.1), we can formally show that $\rho \approx 1$ is the most conservative assumption possible, in the sense that it minimizes the t-statistic for $\beta = 0$. The variable μ_t in (A.1) can be written as a function of the assumed ρ: $\mu_t = x_t - \rho\, x_{t-1}$. Substituting yields

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

As we change ρ, we simply add or subtract x_{t-1} from the second explanatory variable. Doing so affects only the estimate of β, not the residuals v_t or the estimate of γ. Since $\hat{\beta}_{adj}$ is a linear function of ρ, with slope $\hat{\gamma} < 0$, it is minimized at $\rho \approx 1$. If we ignore sampling error in γ, then we can stop here; the standard error of $\hat{\beta}_{adj}$ does not depend on ρ, so the t-statistic is also minimized at $\rho \approx 1$.[13] However, with sampling error in γ, the standard error does depend somewhat on the assumed ρ. The standard error 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:

$$x_{t-1} = c + \delta\, \mu_t + s_t. \tag{A.5}$$

Re-write $\mu_t = \hat{\mu}_t + (\hat{\rho} - \rho)\, x_{t-1}$, where $\hat{\rho}$ and $\hat{\mu}_t$ are the OLS estimates from the AR1 regression. The variance of s_t is:

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

In this equation, all of the variances are sample statistics, not adjusted for degrees of freedom. The second equality uses the fact that $\hat{\mu}_t$ and x_{t-1} are uncorrelated by construction. Re-arranging yields

---

[13] If γ is known, the standard error of β_{adj} is $[\, s_v^2 (X'X)^{-1}_{(2,2)}\,]^{1/2}$, where s_v^2 is the sample variance of v_t (with T – 2 degrees of freedom). Neither v_t nor X depends on the assumed ρ. Empirically, γ can be estimated very precisely and estimation error in γ has almost no affect on the tests. For example, in full-sample regressions with the value-weighted index, $\hat{\gamma}$ is –90.4 with a standard error of 1.1.

$$\sigma_s^2 = \sigma_x^2\left[\frac{\sigma_{\mu,OLS}^2}{\sigma_{\mu,OLS}^2 + (\hat{\rho}-\rho)^2\sigma_x^2}\right]. \tag{A.7}$$

This equation shows that σ_s^2 decreases as $(\hat{\rho}-\rho)^2$ gets larger. For reasonably values of ρ, this effect is small and the t-statistic is approximately linear in ρ (it is dominated by the changes in the numerator and minimized at $\rho \approx 1$). More formally, let $k = T^{1/2}\,\sigma_x\,\sigma_{\mu,OLS} / \hat{\sigma}_v$. The t-statistic from (A.1) is

$$t_{\beta=0} = k\frac{\hat{\beta}-\hat{\gamma}(\hat{\rho}-\rho)}{\left[\sigma_{\mu,OLS}^2 + (\hat{\rho}-\rho)^2\sigma_x^2\right]^{1/2}}. \tag{A.8}$$

The numerator is decreasing in ρ, while the denominator is increasing in $(\hat{\rho}-\rho)^2$. This function has a unique local extremum at $\rho = \hat{\rho} + (\hat{\gamma}\ \sigma_{\mu,OLS}^2 \ /\ \hat{\beta}\ \sigma_x^2)$. Assuming $\hat{\beta} > 0$, this is a global maximum and occurs at some $\rho < \hat{\rho}$. Thus, to show that the t-statistic is minimized at $\rho = 1$, we just need to show that it is higher at $\rho = -1$ than at $\rho = 1$. If the global maximum occurs at $\rho < -1$, the proof is trivial because the derivative of (A.8) is everywhere negative between –1 and 1. Suppose the maximum occurs at $\rho > -1$. As $\rho \to -\infty$, the t-statistic asymptotes to $-k\hat{\gamma} / \sigma_x$ from above. Substituting for k, this implies that the t-statistic at $\rho = -1$ must be greater than $-\hat{\gamma}\ T^{1/2}\ \sigma_{\mu,OLS}\ /\ \hat{\sigma}_v$. This expression is just the t-statistic for testing whether $-\hat{\gamma}$ equals zero (i.e., for testing whether shocks to returns and shocks to DY are related). For applications of interest, this t-statistic will be very large and much greater than $t_{\beta=0}$ assuming $\rho = 1$ (e.g., it is 82.1 in the full-sample VWNY regressions).

References

Ang, Andrew and Geert Bekaert, 2002. Stock return predictability: Is it there?. Working paper (Columbia University, New York).

Blanchard, Olivier and Mark Watson, 1982. Bubbles, rational expectations, and financial markets. In: Paul Watchel, ed., *Crises in the Economic and Financial Structure*. Lexington Books, Lexington, MA, 295-315.

Campbell, John, 1987. Stock returns and the term structure. *Journal of Financial Economics* 18 (2), 373-399.

Campbell, John and Robert Shiller, 1988. The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195-228.

DeJong, David and Charles Whiteman, 1991. The temporal stability of dividends and stock prices: Evidence from the likelihood function. *American Economic Review* 81, 600-617.

Diba, Behzad and Herschel Grossman, 1988. Explosive rational bubbles in stock prices?. *American Economic Review* 78 (3), 520-530.

Fama, Eugene and Kenneth French, 1988. Dividend yields and expected stock returns. *Journal of Financial Economics* 22, 3-25.

Fama, Eugene and Kenneth French, 2002. The equity premium. *Journal of Finance* 57, 637-659.

Flood, Robert and Robert Hodrick, 1986. Asset price volatility, bubbles, and process switching. *Journal of Finance* 41 (4), 831-842.

Flood, Robert and Robert Hodrick, 1990. On testing for speculative bubbles. *Journal of Economic Perspectives* 4 (2), 85-101.

Goetzmann, William and Philippe Jorion, 1993. Testing the predictive power of dividend yields. *Journal of Finance* 48, 663-679.

Hamilton, James and Charles Whiteman, 1985. The observable implications of self-fulfilling expectations. *Journal of Monetary Economics* 16, 353-373.

Hodrick, Robert, 1992. Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies* 5, 357-386.

Kendall, Maurice, 1953. The analysis of economic time series, part I: Prices. *Journal of the Royal Statistical Society* 96, 11-25.

Kendall, Maurice, 1954. Note on bias in the estimation of autocorrelation. *Biometrika* 41, 403-404.

Kothari, S.P. and Jay Shanken, 1997. Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics* 44, 169-203.

Kothari, S.P., Jay Shanken, and Richard Sloan, 1995. Another look at the cross-section of expected stock returns. *Journal of Finance* 50, 185-224.

Lamont, Owen, 1998. Earnings and expected returns. *Journal of Finance* 53, 1563-1587.

Loewenstein, Mark and Gregory Willard, 1999. Rational equilibrium asset-pricing bubbles in continuous trading models, forthcoming in *Journal of Economic Theory*.

Mankiw, N. Gregory and Matthew Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectations models. *Economic Letters* 20, 139-145.

Nelson, Charles and Myung Kim, 1993. Predictable stock returns: The role of small sample bias. *Journal of Finance* 48, 641-661.

Pontiff, Jeffrey and Lawrence Schall, 1998. Book-to-market ratios as predictors of market returns. *Journal of Financial Economics* 49, 141-160.

Stambaugh, Robert, 1986. Bias in regressions with lagged stochastic regressors. Working paper (University of Chicago, Chicago, IL).

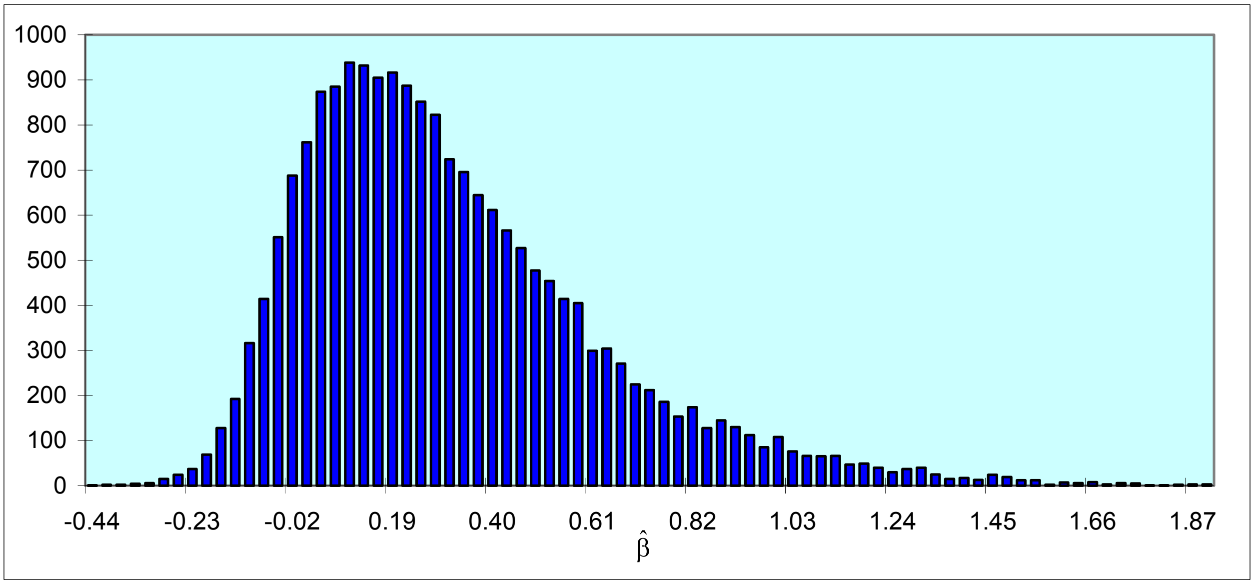
Stambaugh, Robert, 1999. Predictive regressions. *Journal of Financial Economics* 54, 375-421.

Tirole, Jean, 1982. On the possibility of speculation under rational expectations. *Econometrica* 50, 1163-1181.

Tirole, Jean, 1985. Asset bubbles and overlapping generations. *Econometrica* 53, 1071-1100.

West, Kenneth, 1988. Bubbles, fads, and stock price volatility tests: A partial evaluation. Journal of Finance 43, 639-656.

Panel A: Marginal distribution of $\hat{\beta}$



Panel B: Joint distribution of $\hat{\beta}$ and $\hat{\rho}$

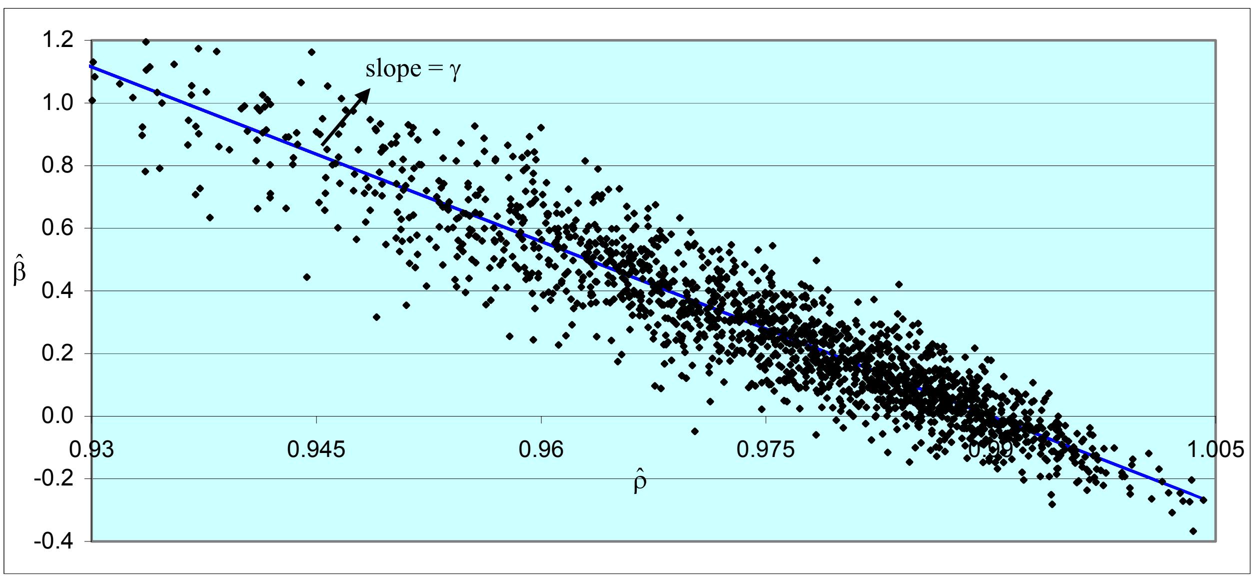


Figure 1
Sampling distribution of $\hat{\beta}$ and $\hat{\rho}$

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 slope coefficients are $\beta = 0$ and $\rho = 0.99$, $cor(\varepsilon, \mu) = -0.92$, $\sigma_\varepsilon = 0.04$, $\sigma_\mu = 0.002$, and $T = 300$.

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 logarithm 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 defore depreciation to market equity.

				Autocorrelation		
Variable	Mean	S.D.	Skew.	ρ_1	ρ_{12}	ρ_{24}
			Returns and dividend yield			
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
			Book-to-market and earnings-price ratio			
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

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 Jan. 1946 – Dec. 2000 (660 months). Observations are monthly. DY is the dividend yield on the value-weighted NYSE index and 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 Tbill 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.

$Log(DY_t) = \phi + \rho \, Log(DY_{t-1}) + \mu_t$

		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

$r_t = \alpha + \beta \, Log(DY_{t-1}) + \varepsilon_t$

		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			

Table 3
Dividend yield and expected returns, 1946 – 1972 and 1973 – 2000

The table reports AR1 regressions for dividend yield and predictive regressions for stock returns for two periods, Jan. 1946 – Dec. 1972 and Jan. 1973 – Dec. 2000 (324 and 336 months, respectively). Observations are monthly. DY is the dividend yield on the value-weighted NYSE index and 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 Tbill 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.

$\text{Log}(DY_t) = \phi + \rho\, \text{Log}(DY_{t-1}) + \mu_t$

		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)
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

$r_t = \alpha + \beta\, \text{Log}(DY_{t-1}) + \varepsilon_t$

		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			

Table 4
Influence of 1995 – 2000

The table reports AR1 regressions for dividend yield and predictive regressions for stock returns for two periods, Jan. 1946 – Dec. 1994 (588 months) and Jan. 1946 – Dec. 2000 (660 months). Observations are monthly. DY is the dividend yield on the value-weighted NYSE index and 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.

Jan. 1946 – Dec. 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)	p-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			

Jan. 1946 – Dec. 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)	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			

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 – Dec. 1994 (379 months) and June 1963 – Dec. 2000 (451 months). Observations are monthly. B/M is the ratio of book equity to market equity on the value-weighted NYSE index. Log(B/M) is the natural logarithm of B/M. EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns equal EWNY and VWNY minus the one-month Tbill rate. Prices and returns come from CRSP, book equity comes 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.

$\text{Log}(B/M_t) = \phi + \rho \, \text{Log}(B/M_{t-1}) + \mu_t$

		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	0.009	-0.013	-0.010	0.972	0.047	0.995	0.006	-0.011	-0.009	0.983	0.047

$r_t = \alpha + \beta \, \text{Log}(B/M_{t-1}) + \varepsilon_t$

		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	1.801	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			
EWNY	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.EWNY	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 – Dec. 1994 (379 months) and June 1963 – Dec. 2000 (451 months). Observations are monthly. E/P is the ratio of operating earnings to market equity on the value-weighted NYSE index. Log(E/P) is the natural logarithm of E/P. EWNY and VWNY are returns on equal- and value-weighted NYSE indexes, respectively. Excess returns equal EWNY and VWNY minus the one-month Tbill rate. Prices and returns come from CRSP, earnings come 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.

$\text{Log}(E/P_t) = \phi + \rho\, \text{Log}(E/P_{t-1}) + \mu_t$

		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

$r_t = \alpha + \beta\, \text{Log}(E/P_{t-1}) + \varepsilon_t$

		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			
EWNY	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.EWNY	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			