

NBER WORKING PAPER SERIES

EFFICIENT TESTS OF STOCK RETURN PREDICTABILITY

John Y. Campbell  
Motohiro Yogo

Working Paper 10026  
<http://www.nber.org/papers/w10026>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2003

We have benefited from comments by Andrew Ang, Geert Bekaert, Markku Lanne, Marcelo Moreira, Robert Shiller, Robert Stambaugh (discussant), James Stock, Mark Watson, and seminar participants at Harvard, the Econometric Society Australasian Meeting 2002, and the NBER Summer Institute 2003. The views expressed herein are those of the authors and are not necessarily those of the National Bureau of Economic Research.

©2003 by John Y. Campbell and Motohiro Yogo. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Efficient Tests of Stock Return Predictability  
John Y. Campbell and Motohiro Yogo  
NBER Working Paper No. 10026  
October 2003  
JEL No. C22, G1

### **ABSTRACT**

Tests of the predictability of stock returns may be invalid when the predictor variable is persistent and its innovations are highly correlated with returns. This paper proposes two methods to deal with the problem. First, we develop a pretest that determines when the conventional t-test is misleading. Second, we develop a new test of predictability that always leads to correct inference and is efficient compared to existing methods. Applying our methods to US data, we find that the conventional t-test is highly misleading for the dividend-price ratio and the smoothed earnings-price ratio. However, we find evidence for predictability using our test, particularly with the earnings-price ratio. We also find evidence for predictability with the short-term interest rate and the long-short yield spread, for which the conventional t-test leads to correct inference.

John Y. Campbell  
Department of Economics  
Harvard University  
Littauer Center 213  
Cambridge, MA 02138  
and NBER  
[john\\_campbell@harvard.edu](mailto:john_campbell@harvard.edu)

Motohiro Yogo  
Department of Economics  
Harvard University  
Littauer Center  
Cambridge, MA 02138  
[yogo@fas.harvard.edu](mailto:yogo@fas.harvard.edu)

# 1 Introduction

Numerous studies in the last two decades have asked whether stock returns can be predicted by financial variables such as the dividend-price ratio, the earnings-price ratio, and various measures of the interest rate. (See for example Keim and Stambaugh (1986), Campbell (1987), Campbell and Shiller (1988), Fama and French (1988, 1989), and Hodrick (1992).) The econometric method used in a typical study is an OLS regression of stock returns onto the lag of the financial variable. The main finding of such regressions is that the  $t$ -statistic is typically greater than two and sometimes greater than three. Using conventional critical values for the  $t$ -test, one would conclude that there is strong evidence for the predictability of returns.

This statistical inference of course relies on first-order asymptotic distribution theory, which implies that the  $t$ -statistic is approximately standard normal in large samples. However, both simulation and analytical studies have shown that the large-sample theory provides a poor approximation to the actual finite-sample distribution of test statistics when the predictor variable is persistent and its innovations are highly correlated with returns (see Mankiw and Shapiro (1986), Elliott and Stock (1994), and Stambaugh (1999)).

To be concrete, suppose the log dividend-price ratio is used to predict returns. Even if we were to know on prior grounds that the dividend-price ratio is stationary, a time series plot (or more formally a unit root test) shows that it is highly persistent, much like a nonstationary process. Since first-order asymptotics fails when the regressor is nonstationary, it provides a poor approximation in finite samples when the regressor is persistent. Elliott and Stock (1994, Table 1) provide Monte Carlo evidence which suggests that the size distortion of the one-sided  $t$ -test is approximately 20 percentage points for plausible parameter values and sample size in the dividend-price ratio regression.<sup>1</sup> They propose an alternative asymptotic framework in which the regressor is modeled as having a local-to-unit root, which provides an accurate approximation to the finite-sample distribution.

These econometric problems have led some recent papers to reexamine the evidence

---

<sup>1</sup>We report their result for the one-sided  $t$ -test at the 10% level when the sample size is 100, the regressor follows an AR(1) with an autoregressive coefficient of 0.975, and the correlation between the innovations to the dependent variable and the regressor is -0.9.

for predictability using tests that have the correct size even if the predictor variable is highly persistent or contains a unit root. Torous, Valkanov, and Yan (2001) develop a test procedure, extending the work of Richardson and Stock (1989) and Cavanagh, Elliott, and Stock (1995), and find evidence for predictability at short horizons but not at long horizons. By testing the stationarity of long-horizon returns, Lanne (2002) concludes that stock returns cannot be predicted by a highly persistent predictor variable. Building on the finite-sample theory of Stambaugh (1999), Lewellen (2003) finds some evidence for predictability with valuation ratios.

A difficulty with understanding the rather large literature on predictability is the sheer variety of test procedures that have been proposed, which have led to different conclusions about the predictability of returns. The main contribution of this paper is to provide an understanding of the various test procedures and their empirical implications within the unifying framework of statistical optimality theory. When the degree of persistence of the predictor variable is known, there is a uniformly most powerful (UMP) test conditional on an ancillary statistic. Although the degree of persistence is not known in practice, this provides a useful benchmark for thinking about the relative power advantages of the various test procedures. In particular, Lewellen's (2003) test is UMP when the predictor variable contains a unit root.

Based on the infeasible UMP test, we propose a new test procedure that is computationally simple (i.e. can be implemented with standard regression output) and has good power (i.e. more powerful than the Bonferroni  $t$ -test of Cavanagh, Elliott, and Stock (1995)). Our test is asymptotically valid under fairly general assumptions about the dynamics of the predictor variable (i.e. a finite-order autoregression with the largest root less than, equal to, or even greater than one) and the distribution of the innovations (i.e. even heteroskedastic). The intuition for our approach is as follows. A regression of stock returns onto a lagged financial variable has low power because stock returns are extremely noisy. If we can eliminate some of this noise, we can increase the power of the test. When the innovations to returns and the predictor variable are correlated, we can subtract off the part of the innovation to the predictor variable that is correlated with returns to obtain a less noisy dependent variable for our regression. Of course, this procedure requires us to measure the innovation to the

predictor variable. When the predictor variable is highly persistent, it is possible to do so in a way that retains power advantages over the conventional regression.

Although tests derived under local-to-unity asymptotics (e.g. Cavanagh, Elliott, and Stock (1995) or the one proposed in this paper) always lead to correct inference, they can be somewhat more difficult to implement than the conventional  $t$ -test. A researcher may therefore be interested in knowing when the conventional  $t$ -test leads to correct inference. We therefore develop a simple pretest based on the confidence interval for the largest autoregressive root of the predictor variable. If the confidence interval indicates that the predictor variable is sufficiently stationary, for a given level of correlation between the innovations to returns and the predictor variable, one can proceed with inference based on the  $t$ -test with conventional critical values.

We apply our methods to annual, quarterly, and monthly US data, looking first at the dividend-price and the smoothed earnings-price ratios. Using the pretest, we find that these valuation ratios are sufficiently persistent for the conventional  $t$ -test to be misleading. Using our test that is robust to the persistence problem, we find that the earnings-price ratio reliably predicts returns at all frequencies in the full sample since 1926. The dividend-price ratio also predicts returns at annual frequency, but we cannot reject the null hypothesis at quarterly and monthly frequencies.

In the sub-sample since 1952, we find that the dividend-price ratio predicts returns at all frequencies if its largest autoregressive root is less than or equal to one. However, since statistical tests do not reject an explosive root for the dividend-price ratio, we have evidence for return predictability only if we are willing to rule out an explosive root based on prior knowledge. This reconciles the “contradictory” findings by Torous, Valkanov, and Yan (2001, Table 3), who report that the dividend-price ratio does not predict monthly returns in the postwar sample, and Lewellen (2003, Table 2), who reports strong evidence for predictability.

Finally, we consider the short-term nominal interest rate and the long-short yield spread as predictor variables in the period since 1952. Our pretest indicates that the conventional  $t$ -test is valid for these interest rate variables since their innovations have low correlation with returns. Using either the conventional  $t$ -test or our more generally valid test procedure, we find strong evidence that these variables predict returns.

The rest of the paper is organized as follows. In Section 2, we derive the UMP test of predictability when the degree of persistence of the predictor variable is known. We then compare its power to that of the conventional  $t$ -test under first-order asymptotics. Although first-order asymptotics is not applicable for persistent predictor variables, the calculations provide intuition for the UMP test in a familiar framework. In Section 3, we compare the power of various tests of predictability under local-to-unity asymptotics and show that a feasible version of the UMP test has good power. We also introduce the pretest for determining when the conventional  $t$ -test leads to correct inference. In Section 4, we apply our test procedure to US equity data and reexamine the empirical evidence for predictability. We reinterpret previous empirical studies within our unifying framework. Section 5 concludes. A separate appendix (Campbell and Yogo 2003) contains tables necessary for implementing the econometric methods in this paper.

## 2 Predictive Regressions

### 2.1 The Regression Model

Let  $r_t$  denote the excess stock return in period  $t$ , and let  $x_{t-1}$  denote a variable observed at  $t-1$  which may have the ability to predict  $r_t$ . For instance,  $x_{t-1}$  may be the log dividend-price ratio at  $t-1$ . The regression model that we consider is

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

$$x_t = \rho x_{t-1} + e_t, \quad (2)$$

for  $t = 1, \dots, T$  with  $x_0 = 0$ .  $\beta$  is the unknown coefficient of interest. We say that the variable  $x_{t-1}$  has the ability to predict returns if  $\beta \neq 0$ . For simplicity, we assume that both  $r_t$  and  $x_t$  have mean zero, so the usual intercept terms do not appear in equations (1) and (2). In addition, we assume that  $(u_t, e_t)'$  is independently and identically distributed (i.i.d.) normal with mean zero and covariance matrix

$$\bar{\Sigma} = \begin{bmatrix} 1 & \delta \\ \delta & 1 \end{bmatrix}. \quad (3)$$

We further assume that the correlation  $\delta \leq 0$  between the innovations is known. (The negativity of  $\delta$  is without loss of generality since the sign of  $\beta$  is unrestricted; redefining the predictor variable as  $-x_t$  flips the signs of both  $\beta$  and  $\delta$ .) We will later relax these assumptions to a more realistic model. For now, this simple model captures the essence of the problem.

In equation (2),  $\rho$  is the unknown degree of persistence in the variable  $x_t$ . If  $|\rho| < 1$  and fixed,  $x_t$  is integrated of order zero, denoted as  $I(0)$ . If  $\rho = 1$ ,  $x_t$  is integrated of order one, denoted as  $I(1)$ . Since  $\beta$  and  $\rho$  are the only unknown parameters in the model, (two times) the joint log likelihood is given by

$$L(\beta, \rho) = -\frac{1}{1 - \delta^2} \sum_{t=1}^T [(r_t - \beta x_{t-1})^2 - 2\delta(r_t - \beta x_{t-1})(x_t - \rho x_{t-1}) + (x_t - \rho x_{t-1})^2], \quad (4)$$

up to an additive constant.

The focus of this paper is the null hypothesis  $\beta = \beta_0$ . We consider two types of alternative hypotheses. The first is the simple alternative  $\beta = \beta_1$ , and the second is the one-sided (composite) alternative  $\beta > \beta_0$ . The hypothesis testing problem is complicated by the fact that  $\rho$  is an unknown nuisance parameter.

## 2.2 Likelihood Ratio Test

One way to test the hypothesis of interest in the presence of the nuisance parameter  $\rho$  is through the likelihood ratio test (LRT). Let  $\hat{\beta}$  be the OLS estimator of  $\beta$ , and let

$$t(\beta_0) = \frac{\hat{\beta} - \beta_0}{(\sum_{t=1}^T x_{t-1}^2)^{-1/2}} \quad (5)$$

be the associated  $t$ -statistic. The LRT rejects the null if

$$\max_{\beta, \rho} L(\beta, \rho) - \max_{\rho} L(\beta_0, \rho) = t(\beta_0)^2 > C, \quad (6)$$

for some constant  $C$ . (With a slight abuse of notation, we use  $C$  to denote a generic constant throughout the paper.) In other words, the LRT corresponds to the  $t$ -test.

Note that we would obtain the same test (6) starting from the marginal likelihood  $L(\beta) = -\sum_{t=1}^T (r_t - \beta x_{t-1})^2$ . The LRT can thus be interpreted as a test that ignores information

contained in equation (2) of the regression model. Although the LRT is not derived from statistical optimality theory, it has desirable large-sample properties when  $x_t$  is  $I(0)$  (see Cox and Hinkley (1974)). For instance, the  $t$ -statistic is asymptotically pivotal, that is, its asymptotic distribution does not depend on the nuisance parameter  $\rho$ . The  $t$ -test is therefore a reasonable solution to the hypothesis testing problem when  $x_t$  is  $I(0)$  and  $\rho$  is unknown.

## 2.3 Optimal Test When $\rho$ is Known

The problem with the nuisance parameter can also be resolved if  $\rho$  were known *a priori*. Since  $\beta$  is then the only unknown parameter in the likelihood function (4), the Neyman-Pearson Lemma implies that the most powerful test against the simple alternative rejects the null if

$$\begin{aligned} (1 - \delta^2)(L(\beta_1, \rho) - L(\beta_0, \rho)) &= 2(\beta_1 - \beta_0) \sum_{t=1}^T x_{t-1} [r_t - \delta(x_t - \rho x_{t-1})] \\ &\quad - (\beta_1^2 - \beta_0^2) \sum_{t=1}^T x_{t-1}^2 > C. \end{aligned} \quad (7)$$

Since the optimal test statistic is a weighted sum of two minimal sufficient statistics with the weights depending on the alternative  $\beta_1$ , there is no UMP test.

However, the second statistic  $\sum_{t=1}^T x_{t-1}^2$  is ancillary (i.e. its distribution does not depend on  $\beta$ ). Hence, it is natural to restrict ourselves to tests that condition on the ancillary statistic. The optimal *conditional* test rejects the null if

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^T x_{t-1} [r_t - \beta_0 x_{t-1} - \delta(x_t - \rho x_{t-1})]}{(1 - \delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^2)^{1/2}} > C. \quad (8)$$

Since the test takes the same form for all  $\beta_1 > \beta_0$ , it is UMP against one-sided alternatives when  $\rho$  is known. For simplicity, we will refer to this (infeasible) test as the  $Q$ -test.

When  $\beta_0 = 0$ ,  $Q(\beta_0, \rho)$  is the  $t$ -statistic that results from regressing  $r_t - \delta(x_t - \rho x_{t-1})$  onto  $x_{t-1}$ . It collapses to the conventional  $t$ -statistic (5) when  $\delta = 0$ . Since  $e_t = x_t - \rho x_{t-1}$ , knowledge of  $\rho$  allows us to subtract off the part of innovation to returns that is correlated with the innovation to the predictor variable, resulting in a more powerful test. If we let  $\hat{\rho}$  denote the OLS estimator of  $\rho$ , then the  $Q$ -statistic can be written as

$$Q(\beta_0, \rho) = \frac{(\hat{\beta} - \beta_0) - \delta(\hat{\rho} - \rho)}{(1 - \delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^2)^{-1/2}}. \quad (9)$$



Drawing on the work of Stambaugh (1999), Lewellen (2003) motivates the statistic by interpreting the term  $\delta(\hat{\rho} - \rho)$  as the “finite-sample bias” of the OLS estimator. Assuming that  $\rho \leq 1$ , Lewellen tests the predictability of returns using the statistic  $Q(\beta_0, 1)$ .

## 2.4 Power under First-Order Asymptotics

We now derive the power functions of the  $t$ -test and the  $Q$ -test under first-order asymptotics to illustrate the power gains that would result from incorporating knowledge of the persistence parameter  $\rho$ . When  $x_t$  is  $I(0)$ , the OLS estimator  $\hat{\beta}$  is  $\sqrt{T}$ -consistent. Hence, any reasonable test, such as the conventional  $t$ -test, rejects alternatives that are a fixed distance from the null with probability one as the sample size becomes arbitrarily large. In practice, however, we have a finite sample and are interested in the relative efficiency of test procedures. A natural way to evaluate the power of tests in finite samples is to consider their ability to reject local alternatives.<sup>2</sup> Formally, we consider a sequence of alternatives of the form  $\beta = \beta_0 + b/\sqrt{T}$  for some fixed constant  $b$ .

Let  $\Phi(z)$  denote one minus the standard normal cumulative distribution function, and let  $z_\alpha$  denote the upper  $\alpha$ -quantile of that distribution. Under first-order asymptotics, the probability that the  $t$ -test rejects a local alternative  $b$  is

$$\pi_t(b) = \Phi(z_\alpha - |b|\sigma_x), \quad (10)$$

where  $\sigma_x^2 = \mathbf{E}[x_t^2] = 1/(1 - \rho^2)$ . Similarly, the power function of the  $Q$ -test is

$$\pi_Q(b) = \Phi\left(z_\alpha - \frac{|b|\sigma_x}{(1 - \delta^2)^{1/2}}\right). \quad (11)$$

Since the  $Q$ -test is UMP against one-sided alternatives conditional on the ancillary statistic  $T^{-1} \sum_{t=1}^T x_{t-1}^2$ ,  $\pi_Q(b)$  is the power envelope for conditional tests when  $\rho$  is known. Moreover, it is asymptotically the power envelope for all (including unconditional) tests since the ancillary statistic has a degenerate asymptotic distribution.

Figure 1 shows the power functions for various combinations of  $\rho$  (0.99 and 0.75) and  $\delta$  (-0.95 and -0.75). These values are chosen to correspond to the relevant region of the parameter space when the predictor variable is a valuation ratio (i.e. the log dividend-price

---

<sup>2</sup>See Lehmann (1999, Chapter 3) for a textbook treatment of local alternatives and relative efficiency.

ratio or the log earnings-price ratio). As expected, the power of the  $Q$ -test dominates that of the  $t$ -test. A comparison of (10) and (11) shows that the power gain arises from  $\delta^2 \neq 0$  and is increasing in the degree of correlation. Intuitively, when  $\rho$  is known, the innovation  $e_t = x_t - \rho x_{t-1}$  is known as well. Then by subtracting off the portion of the innovation  $u_t$  that is correlated with  $e_t$  (i.e.  $\delta e_t$ ), the  $Q$ -test is able to gain power through the reduction in noise. When the predictor variable is a valuation ratio, the power gain from using the  $Q$ -test is especially large since the innovations to returns and the valuation ratio are highly correlated through the stock price. In practice, the  $Q$ -test is infeasible because  $\rho$  is unknown, so unfortunately, the large power gains over the  $t$ -test cannot be realized.

## 2.5 Generalizing the Regression Model

The regression model (1)–(2) in which the  $Q$ -test (8) is UMP, conditional on an ancillary statistic, is restrictive. In this section, we show that a generalization of the  $Q$ -test is UMP invariant, conditional on an ancillary statistic, in a more realistic and empirically relevant model.

Consider the regression model

$$r_t = \alpha + \beta x_{t-1} + u_t, \quad (12)$$

$$x_t = \gamma + \rho x_{t-1} + e_t, \quad (13)$$

for  $t = 1, \dots, T$ . The model now includes intercept terms assumed away in the simplified model (1)–(2), which account for the mean of returns and the predictor variable. Their magnitudes depend on the units in which the variables are measured. For instance, there is an arbitrary scaling factor involved in computing the dividend-price ratio, which results in an arbitrary constant shifting the level of the log dividend-price ratio. Since we do not want inference to depend on the units in which the variables are measured, it is natural to restrict ourselves to tests that are invariant to translations in  $\alpha$  and  $\gamma$  (see Lehmann (1986, Chapter 6)). That is, we only consider test statistics whose values do not change with additive shifts in  $r_t$  and  $x_t$ .

Now suppose  $\rho$  is known and the following distributional assumptions hold.

**Assumption 1 (Normality).**  $w_t = (u_t, e_t)'$  is independently distributed  $\mathbf{N}(0, \Sigma)$ , where

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{bmatrix}$$

is known.  $x_0$  is fixed and known.

Let  $x_{t-1}^\mu = x_{t-1} - T^{-1} \sum_{t=1}^T x_{t-1}$  be the de-measured predictor variable. By the same argument as that in Section 2.3, the test based on the statistic

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^T x_{t-1}^\mu [r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e^2} (x_t - \rho x_{t-1})]}{\sigma_u (1 - \delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}. \quad (14)$$

is UMP conditional on the ancillary statistic  $\sum_{t=1}^T x_{t-1}^{\mu 2}$ . In other words, the statistic (14) is a generalization of (8) to the model (12)–(13). The asymptotic power function for the  $Q$ -test is given by (11), when the variances are normalized as  $\sigma_u^2 = \sigma_e^2 = 1$ . Moreover, it corresponds to the power envelope for tests that use knowledge of the the autoregressive root  $\rho$ .

### 3 Inference with a Persistent Regressor

Figure 2 is a time series plot of the log dividend-price ratio for the CRSP NYSE/AMEX value-weighted index and the log smoothed earnings-price ratio for the S&P 500 index at quarterly frequency. Following Campbell and Shiller (1988), earnings are smoothed by taking a backwards moving average over ten years. Both valuation ratios are persistent and even appear to be nonstationary, especially toward the end of the sample period. The 95% confidence intervals for  $\rho$  are  $[0.957, 1.007]$  and  $[0.939, 1.000]$  for the dividend-price ratio and the earnings-price ratio, respectively.

The persistence of financial variables typically used to predict returns has important implications for inference about predictability. Even if the predictor variable is  $I(0)$ , first-order asymptotics can be a poor approximation in finite samples when  $\rho$  is close to one because of the discontinuity in the asymptotic distribution at  $\rho = 1$ . Inference based on first-order asymptotics may therefore be invalid due to size distortions. The solution is to base

inference on more accurate approximations to the actual (unknown) sampling distribution of test statistics. There are two main approaches that have been used in the literature.

The first approach is the exact finite-sample theory under the assumption of normality (i.e. Assumption 1). This is the approach taken by Evans and Savin (1981, 1984) for autoregression and Stambaugh (1999) for predictive regressions. The second approach is local-to-unity asymptotics, which has been applied successfully to approximate the finite-sample behavior of persistent time series in the unit root testing literature. (See Stock (1994) for a survey and references.) Local-to-unity asymptotics has been applied to the present context of predictive regressions by Elliott and Stock (1994), who derived the asymptotic distribution of the  $t$ -statistic. This has been extended to long-horizon  $t$ -tests by Torous, Valkanov, and Yan (2001). The advantage of local-to-unity asymptotics, over the exact Gaussian theory, is that it allows for a wide variety of distributions for the innovations, including short-run dynamics and heteroskedasticity. Hence, this is the approach that we follow in this paper.

### 3.1 Local-to-Unity Asymptotics

The AR(1) model for the predictor variable (13) is restrictive since it does not allow for short-run dynamics. Let  $L$  be the lag operator. We generalize the model as

$$x_t = \gamma + \rho x_{t-1} + v_t, \quad (15)$$

$$b(L)v_t = e_t, \quad (16)$$

where  $b(L) = \sum_{i=0}^{p-1} b_i L^i$  with  $b_0 = 1$  and  $b(1) \neq 0$ . All the roots of  $b(L)$  are assumed to be fixed and less than one in absolute value. Equations (15) and (16) together imply that

$$\Delta x_t = \tau + \theta x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + e_t. \quad (17)$$

Hence, the dynamics of the predictor variable is captured by an AR(p), which is written here in the augmented Dickey-Fuller form.

To apply the asymptotic theory, we assume that the sequence of innovations satisfies the following fairly weak distributional assumptions.

**Assumption 2 (Martingale Difference Sequence).** *Let  $\mathcal{F}_t = \{w_s | s \leq t\}$  be the filtration generated by the process  $w_t = (u_t, e_t)'$ . Then*

1.  $\mathbf{E}[w_t | \mathcal{F}_{t-1}] = 0$ ,
2.  $\mathbf{E}[w_t w_t'] = \Sigma$ ,
3.  $\sup_t \mathbf{E}[u_t^4] < \infty$ ,  $\sup_t \mathbf{E}[e_t^4] < \infty$ , and  $\mathbf{E}[x_0^2] < \infty$ .

In other words,  $w_t$  is a martingale difference sequence with finite fourth moments. The assumption allows the sequence of innovations to be conditionally heteroskedastic as long as it is covariance stationary (i.e. unconditionally homoskedastic). Assumption 1 is a special case when the innovations are i.i.d. normal and the covariance matrix  $\Sigma$  is known.

Local-to-unity asymptotics is an asymptotic framework where the the largest autoregressive root is modeled as  $\rho = 1 + c/T$  with  $c$  a fixed constant. Within this framework, the asymptotic distribution theory is not discontinuous when  $x_t$  is  $I(1)$  (i.e.  $c = 0$ ). This device also allows  $x_t$  to be stationary but nearly integrated (i.e.  $c < 0$ ) or even explosive (i.e.  $c > 0$ ). For the rest of the paper, we assume that the true processes for excess returns and the predictor variable are (12) and (15), respectively, where  $c = T(\rho - 1)$  is fixed as  $T$  becomes arbitrarily large. Without loss of generality, we continue to assume that  $\delta = \sigma_{ue}/(\sigma_u \sigma_e) \leq 0$ .

An important feature of the nearly integrated case is that sample moments (e.g. mean and variance) of the process  $x_t$  are not well defined. However, when appropriately scaled, these objects converge to functionals of a diffusion process. Let  $(W_u(s), W_e(s))'$  be a two-dimensional Wiener process with covariance matrix (3). Let  $J_c(s)$  be the diffusion process defined by the stochastic differential equation  $dJ_c(s) = cJ_c(s)ds + dW_e(s)$  with initial condition  $J_c(0) = 0$ . Let  $J_c^\mu(s) = J_c(s) - \int J_c(r)dr$ , where integration is over  $[0, 1]$  unless otherwise noted. Let  $\Rightarrow$  denote weak convergence in the space  $D[0, 1]$  of cadlag functions (see Billingsley (1999, Chapter 3)). Collecting results from Phillips (1987, Lemma 1) and Cavanagh, Elliott, and Stock (1995), we have the following useful lemma.

**Lemma 1 (Weak Convergence).** *Suppose Assumption 2 holds. The following limits hold jointly.*

1.  $T^{-3/2} \sum_{t=1}^T x_t^\mu \Rightarrow \omega \int J_c^\mu(s)ds$ ,

$$2. \quad T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2} \Rightarrow \omega^2 \int J_c^\mu(s)^2 ds,$$

$$3. \quad T^{-1} \sum_{t=1}^T x_{t-1}^\mu v_t \Rightarrow \omega^2 \int J_c^\mu(s) dW_e(s) + \frac{1}{2}(\omega^2 - \sigma_v^2),$$

$$4. \quad T^{-1} \sum_{t=1}^T x_{t-1}^\mu u_t \Rightarrow \sigma_u \omega \int J_c^\mu(s) dW_u(s),$$

where  $\omega = \sigma_e/b(1)$  and  $\sigma_v^2 = \mathbf{E}[v_t^2]$ .

Under first-order asymptotics, the  $t$ -statistic from regression (12) has a standard normal asymptotic distribution. Under local-to-unity asymptotics, the  $t$ -statistic has the null distribution

$$t(\beta_0) \Rightarrow \delta \frac{\tau_c}{\kappa_c} + (1 - \delta^2)^{1/2} Z, \quad (18)$$

where  $\kappa_c = (\int J_c^\mu(s)^2 ds)^{1/2}$ ,  $\tau_c = \int J_c^\mu(s) dW_e(s)$ , and  $Z$  is a standard normal random variable independent of  $(W_e(s), J_c(s))$  (see Elliott and Stock (1994)). Note that the  $t$ -statistic is no longer asymptotically pivotal. That is, its asymptotic distribution depends on an unknown nuisance parameter  $c$  through the random variable  $\tau_c/\kappa_c$ , which makes the test infeasible.

When the predictor variable is an AR(1) (i.e. model (13)), the  $Q$ -statistic (14) has a standard normal asymptotic distribution under the null. Under the more general model (15) which allows for higher-order autocorrelation, the statistic (14) is not asymptotically pivotal. However, a suitably modified statistic

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^T x_{t-1}^\mu [r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e \omega} (x_t - \rho x_{t-1})] + \frac{T}{2} \frac{\sigma_{ue}}{\sigma_e \omega} (\omega^2 - \sigma_v^2)}{\sigma_u (1 - \delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}} \quad (19)$$

has a standard normal asymptotic distribution (see the Appendix for details). In the absence of short-run dynamics (i.e.  $b(1) = 1$  so  $\omega^2 = \sigma_v^2 = \sigma_e^2$ ), the  $Q$ -statistic reduces to (14). The correction term involving  $(\omega^2 - \sigma_v^2)$  is analogous to the correction of the Dickey-Fuller (1981) test by Phillips and Perron (1988).

Although the  $Q$ -statistic is asymptotically pivotal, the test is infeasible since it requires knowledge of  $\rho$  (or equivalently  $c$ ) to compute the test statistic. Even if  $\rho$  were known, the statistic (19) also requires knowledge of the nuisance parameters  $\Sigma$ ,  $\omega^2$ , and  $\sigma_v^2$ . However, a feasible version of the statistic that replaces these nuisance parameters with consistent estimators has the same asymptotic distribution. Therefore, there is no loss of generality in assuming knowledge of these parameters for the purposes of asymptotic theory.

### 3.2 Relation to First-Order Asymptotics and a Simple Pretest

In this section, we first discuss the relationship between first-order and local-to-unity asymptotics. We then develop a simple pretest to determine whether inference based on first-order asymptotics is reliable.

In general, the asymptotic distribution of the  $t$ -statistic (18) is nonstandard because of its dependence on  $\tau_c/\kappa_c$ . However, the  $t$ -statistic is standard normal in the special case  $\delta = 0$ . The  $t$ -statistic should therefore be approximately normal when  $\delta \approx 0$ . Likewise, the  $t$ -statistic should be approximately normal when  $c \ll 0$  because first-order asymptotics is a satisfactory approximation when the predictor variable is stationary. Formally, Phillips (1987, Theorem 2) shows that  $\tau_c/\kappa_c \Rightarrow \tilde{Z}$  as  $c \rightarrow -\infty$ , where  $\tilde{Z}$  is a standard normal random variable independent of  $Z$ .

Figure 3 is a plot of the actual size of the nominal 5% one-sided  $t$ -test as a function of  $c$  and  $\delta$ . In other words, we plot

$$p(c, \delta; \alpha) = \Pr \left( \delta \frac{\tau_c}{\kappa_c} + (1 - \delta^2)^{1/2} Z > z_\alpha \right), \quad (20)$$

where  $\alpha = 0.05$ . The  $t$ -test that uses conventional critical values has approximately the correct size when  $\delta$  is small in absolute value or  $c$  is large in absolute value.<sup>3</sup> The size distortion of the  $t$ -test peaks when  $\delta = -1$  and  $c \approx 1$ . The size distortion arises from the fact that the distribution of  $\tau_c/\kappa_c$  is skewed to the left, which causes the distribution of the  $t$ -statistic to be skewed to the right when  $\delta < 0$ . This causes a right-tailed  $t$ -test that uses conventional critical values to over-reject, and a left-tailed test to under-reject. When the predictor variable is a valuation ratio (e.g. the dividend-price ratio),  $\delta \approx -1$  and the hypothesis of interest is  $\beta = 0$  against the alternative  $\beta > 0$ . Thus we may worry that the evidence for predictability is a consequence of size distortion.

In Table 1, we tabulate the values of  $c \in (c_{\min}, c_{\max})$  for which the size of the right-tailed  $t$ -test exceeds 7.5%, for selected values of  $\delta$ . For instance, when  $\delta = -0.95$ , the nominal 5%  $t$ -test has actual size greater than 7.5% if  $c \in (-79.318, 8.326)$ . The table can be used to construct a pretest to determine whether inference based on the conventional  $t$ -test is

---

<sup>3</sup>The fact that the  $t$ -statistic is approximately normal for  $c \gg 0$  corresponds to asymptotic results for explosive AR(1) with Gaussian errors. See Phillips (1987) for a discussion.

sufficiently reliable. Suppose a researcher is willing to tolerate an actual size of up to 7.5% for a nominal 5% test of predictability. Let  $\Theta = \{c, \delta | p(c, \delta; 0.05) > 0.075\}$ . Then the goal is to test

$$H_0 : \{c, \delta\} \in \Theta$$

$$H_1 : \{c, \delta\} \notin \Theta.$$

To test this hypothesis, we first construct a  $100(1 - \alpha_1)\%$  confidence interval for  $c$ , denoted as  $C_c(\alpha_1)$ . We then estimate  $\delta$  using the residuals from regressions (12) and (17). We reject the null hypothesis if  $C_c(\alpha_1) \cap (c_{\min}, c_{\max}) = \emptyset$ , where  $(c_{\min}, c_{\max})$  is taken from Table 1 using the estimated correlation  $\hat{\delta}$ . That is, we reject the null if the confidence interval for  $c$  lies strictly below (or above) the region of the parameter space  $(c_{\min}, c_{\max})$  where size distortion is large. As emphasized by Elliott and Stock (1994), the rejection of the unit root hypothesis  $c = 0$  is not sufficient to assure that the size distortion is acceptably small. Asymptotically, this pretest has size  $\alpha_1$ .

Since there is no UMP test for an autoregressive unit root (see Elliott, Rothenberg, and Stock (1996)), there is no uniformly most accurate confidence interval for  $c$ . However, as discussed in Elliott and Stock (2001), a relatively accurate confidence interval can be constructed by using a relatively efficient unit root test. In our empirical application, we therefore construct the confidence interval for  $c$  by applying Stock's (1991) method of confidence belts to the DF-GLS test (Elliott, Rothenberg, and Stock 1996). A lookup table for the confidence interval for  $c$ , given the value of the DF-GLS statistic, is provided in Campbell and Yogo (2003, Table 1).

### 3.3 Feasible Tests of Predictability

As discussed in Section 3.1, both the  $t$ -test and the  $Q$ -test are infeasible since the procedures depend on an unknown nuisance parameter  $c$ , which cannot be estimated consistently. Intuitively, the degree of persistence, controlled by the parameter  $c$ , influences the distribution of test statistics that depend on the persistent predictor variable. This must be accounted for by adjusting either the critical values of the test (e.g.  $t$ -test) or the value of the test statistic itself (e.g.  $Q$ -test). Cavanagh, Elliott, and Stock (1995) discuss several (sup-bound,



Bonferroni, and Scheffe-type) methods of making tests that depend on  $c$  feasible. Here, we will discuss the Bonferroni method.

To construct a Bonferroni confidence interval, we first construct a  $100(1 - \alpha_1)\%$  confidence interval for  $\rho$ , denoted as  $C_\rho(\alpha_1)$ . (We parameterize the degree of persistence by  $\rho$  rather than  $c$  since this is the more natural choice in the following.) Then for each value of  $\rho$  in the confidence interval, we construct a  $100(1 - \alpha_2)\%$  confidence interval for  $\beta$  given  $\rho$ , denoted as  $C_{\beta|\rho}(\alpha_2)$ . A confidence interval that does not depend on  $\rho$  can be obtained by

$$C_\beta(\alpha) = \bigcup_{\rho \in C_\rho(\alpha_1)} C_{\beta|\rho}(\alpha_2).$$

By Bonferroni's inequality, this confidence interval has coverage of at least  $100(1 - \alpha)\%$ , where  $\alpha = \alpha_1 + \alpha_2$ .

This approach is conservative in the sense that the actual coverage rate of  $C_\beta(\alpha)$  can be greater than  $100(1 - \alpha)\%$ . This can be seen from the equality

$$\begin{aligned} \Pr(\beta \notin C_\beta(\alpha)) &= \Pr(\beta \notin C_\beta(\alpha) | \rho \in C_\rho(\alpha_1)) \Pr(\rho \in C_\rho(\alpha_1)) \\ &\quad + \Pr(\beta \notin C_\beta(\alpha) | \rho \notin C_\rho(\alpha_1)) \Pr(\rho \notin C_\rho(\alpha_1)). \end{aligned}$$

Since  $\Pr(\beta \notin C_\beta(\alpha) | \rho \notin C_\rho(\alpha_1))$  is unknown, the Bonferroni method bounds it by one as the worst case. In addition, the inequality  $\Pr(\beta \notin C_\beta(\alpha) | \rho \in C_\rho(\alpha_1)) \leq \alpha_2$  is strict unless the conditional confidence intervals  $C_{\beta|\rho}(\alpha_2)$  do not depend on  $\rho$ . Because these worst case conditions are unlikely to hold in practice, the inequality

$$\Pr(\beta \notin C_\beta(\alpha)) \leq \alpha_2(1 - \alpha_1) + \alpha_1 \leq \alpha$$

is likely to be strict, resulting in a conservative confidence interval.

To implement the Bonferroni confidence interval in practice, Cavanagh, Elliott, and Stock suggest inverting the augmented Dickey-Fuller  $t$ -statistic to first construct  $C_\rho(\alpha_1)$ . They then suggest inverting the conventional  $t$ -statistic for testing  $\beta$ , using the appropriate critical values based on its asymptotic distribution (18). The two  $t$ -statistics are correlated, which tends to increase the coverage rate of the confidence interval. Cavanagh, Elliott, and Stock suggest adjusting  $\alpha_1$  and  $\alpha_2$  to achieve an exact test of the desired significance level. Torous, Valkanov, and Yan (2001) have applied this method to test predictability in US data.

A natural question that arises is whether there is a more efficient method of constructing the Bonferroni confidence interval. Since there is no UMP test for an autoregressive unit root, there is no uniformly most accurate confidence interval for  $\rho$ . However, the DF-GLS test is more powerful than the augmented Dickey-Fuller test. Hence, we invert the DF-GLS statistic to obtain a tighter confidence interval for  $\rho$ .

In addition, we know that the  $Q$ -test is a more powerful test of  $\beta$  given  $\rho$  than the  $t$ -test. In fact, it is UMP conditional on an ancillary statistic when  $\rho$  is known. We can therefore obtain a more accurate confidence interval  $C_{\beta|\rho}(\alpha_2)$  by inverting the  $Q$ -test. Because the statistic (19) is asymptotically normal under the null, an equal-tailed  $\alpha_2$ -level confidence interval is simply  $C_{\beta|\rho}(\alpha_2) = [\underline{\beta}(\rho, \alpha_2), \bar{\beta}(\rho, \alpha_2)]$ , where

$$\hat{\beta}(\rho) = \frac{\sum_{t=1}^T x_{t-1}^\mu [r_t - \frac{\sigma_{ue}}{\sigma_e \omega} (x_t - \rho x_{t-1})] + \frac{T}{2} \frac{\sigma_{ue}}{\sigma_e \omega} (\omega^2 - \sigma_v^2)}{\sum_{t=1}^T x_{t-1}^{\mu 2}}, \quad (21)$$

$$\underline{\beta}(\rho, \alpha_2) = \hat{\beta}(\rho) - z_{\alpha_2/2} \sigma_u \left( \frac{1 - \delta^2}{\sum_{t=1}^T x_{t-1}^{\mu 2}} \right)^{1/2}, \quad (22)$$

$$\bar{\beta}(\rho, \alpha_2) = \hat{\beta}(\rho) + z_{\alpha_2/2} \sigma_u \left( \frac{1 - \delta^2}{\sum_{t=1}^T x_{t-1}^{\mu 2}} \right)^{1/2}. \quad (23)$$

Let  $C_\rho(\alpha_1) = [\underline{\rho}(\underline{\alpha}_1), \bar{\rho}(\bar{\alpha}_1)]$  denote the confidence interval for  $\rho$ , where  $\underline{\alpha}_1 = \Pr(\rho < \underline{\rho}(\underline{\alpha}_1))$ ,  $\bar{\alpha}_1 = \Pr(\rho > \bar{\rho}(\bar{\alpha}_1))$ , and  $\alpha_1 = \underline{\alpha}_1 + \bar{\alpha}_1$ . Then the Bonferroni confidence interval is given by

$$C_\beta(\alpha) = [\underline{\beta}(\bar{\rho}(\bar{\alpha}_1), \alpha_2), \bar{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2)]. \quad (24)$$

Hence, we have a closed-form expression for the confidence interval of  $\beta$  that is easy to compute.

As discussed above, the Bonferroni confidence interval can be quite conservative. As suggested by Cavanagh, Elliott, and Stock, the significance levels  $\alpha_1$  and  $\alpha_2$  can be adjusted to achieve a test of desired significance level  $\tilde{\alpha} \leq \alpha$ . To do so, we first fix  $\alpha_2$ . Then for each  $\delta$ , we numerically search over a grid for  $c$  to find the  $\bar{\alpha}_1$  such that

$$\Pr(\underline{\beta}(\bar{\rho}(\bar{\alpha}_1), \alpha_2) > \beta) \leq \tilde{\alpha}/2, \quad (25)$$

with equality at some  $c$ . We then repeat the same procedure for  $\underline{\alpha}_1$  and

$$\Pr(\bar{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2) < \beta) \leq \tilde{\alpha}/2. \quad (26)$$

In Table 2, we report the values of  $\underline{\alpha}_1$  and  $\overline{\alpha}_1$  for selected values of  $\delta$  when  $\tilde{\alpha} = \alpha_2 = 0.10$ , computed over the grid  $c \in [-50, 5]$ . The table can be used to construct a 10% Bonferroni confidence interval for  $\beta$  (equivalently, a 5% one-sided  $Q$ -test for predictability). Note that  $\underline{\alpha}_1$  and  $\overline{\alpha}_1$  are increasing in  $\delta$ , so the Bonferroni inequality has more slack and the unadjusted Bonferroni test is more conservative the smaller is  $\delta$  in absolute value. In order to implement the Bonferroni test using Table 2, one needs the confidence belts for the DF-GLS statistic. In Campbell and Yogo (2003, Tables 2–11), we provide lookup tables which report the appropriate confidence interval for  $c$ ,  $C_c(\alpha_1) = [\underline{c}(\underline{\alpha}_1), \overline{c}(\overline{\alpha}_1)]$ , given the values of the DF-GLS statistic and  $\delta$ . Then the confidence interval  $C_\rho(\alpha_1) = 1 + C_c(\alpha_1)/T$  for  $\rho$  results in a 10% Bonferroni confidence interval for  $\beta$ .

Our computational results indicate that in general the inequalities (25) and (26) are close to equalities when  $c$  is large and more slack when  $c$  is small. For right-tailed tests, the probability (25) can be as small as 4.0% for some values of  $c$  and  $\delta$ . For left-tailed tests, the probability (26) can be as small as 1.2%. This suggests that even the adjusted Bonferroni  $Q$ -test is conservative (i.e. undersized) when  $c < 5$ . The assumption that the predictor variable is never explosive (i.e.  $c \leq 0$ ) allows us to further tighten the Bonferroni confidence interval. In our judgment, however, the magnitude of the resulting power gain is not sufficient to justify the loss of robustness against explosive roots.

## 3.4 Power under Local-to-Unity Asymptotics

### 3.4.1 Infeasible Tests

In this section, we first examine the power of the  $t$ -test and  $Q$ -test under local-to-unity asymptotics. Although these tests assume knowledge of  $c$  and are thus infeasible, their power provide benchmarks for assessing the power of feasible tests. When the predictor variable contains a local-to-unit root, OLS estimators  $\hat{\beta}$  and  $\hat{\rho}$  are consistent at the rate  $T$ , rather than  $\sqrt{T}$ . We therefore consider a sequence of alternatives of the form  $\beta = \beta_0 + b/T$  for some fixed constant  $b$ . Details on the computation of the power functions are in the Appendix.

Figure 4 plots the power functions for the  $t$ -test (using the appropriate critical value that

depends on  $c$ ) and the  $Q$ -test. Under local-to-unity asymptotics, power functions are not symmetric in  $b$ . We only report the power for right-tailed tests (i.e.  $b > 0$ ) since this is the region where the conventional  $t$ -test is size distorted (recall the discussion in Section 3.2). The results, however, are qualitatively similar for left-tailed tests. We consider various combinations of  $c$  (-2 and -20) and  $\delta$  (-0.95 and -0.75), which are in the relevant region of the parameter space when the predictor variable is a valuation ratio. The nuisance parameters are normalized as  $\sigma_u^2 = \omega^2 = 1$ .

As expected, the power function for the  $Q$ -test dominates that for the  $t$ -test. The difference is especially large when  $\delta = -0.95$ . When the correlation between the innovations is large, there are large power gains from subtracting off the part of the innovation to returns that is correlated with the innovation to the predictor variable. These results confirm analogous calculations under first-order asymptotics, reported in Figure 1.

To assess the importance of the power gain, we compute the Pitman efficiency, which is the ratio of the sample sizes at which two tests achieve the same level of power (e.g. 50%) along a sequence of local alternatives. Consider the case  $c = -2$  and  $\delta = -0.95$ . To compute the Pitman efficiency of the  $t$ -test relative to the  $Q$ -test, note first that the  $t$ -test achieves 50% power when  $b = 4.8$ . On the other hand, the  $Q$ -test achieves 50% power when  $b = 1.8$ . Following the discussion in Stock (1994, p. 2775), the Pitman efficiency of the  $t$ -test relative to the  $Q$ -test is  $4.8/1.8 \approx 2.7$ . This means that to achieve 50% power, the  $t$ -test asymptotically requires 170% more observations than the  $Q$ -test.

As was the case under first-order asymptotics, the power function for the  $Q$ -test corresponds to the Gaussian power envelope for conditional tests under model (13) and Assumption 1. Unlike the case for first-order asymptotics, however, the power function does not correspond to the power envelope for all (including unconditional) tests. This is because the ancillary statistic  $T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2}$  has a non-degenerate asymptotic distribution under local-to-unity asymptotics (see Lemma 1). The optimal test statistic is therefore a weighted sum of two statistics, where the weights depend on the alternative  $b$ . Consequently, there is no UMP test against a one-sided alternative. Although not reported in Figure 4, our calculations show that the power of the  $Q$ -test comes very close (essentially tangent) to the power envelope.

When the innovations are not normal, there are in principle tests that more efficient than the  $Q$ -test. However, the  $Q$ -test is asymptotically more efficient than the  $t$ -test even if the innovations are non-normal (i.e. Assumption 2). This illustrates the fact that the Gaussian likelihood function and the Neyman-Pearson Lemma are useful tools for deriving efficient tests even if the error distribution is unknown.

### 3.4.2 Feasible Tests

We now analyze the power properties of several feasible tests that have been proposed. Figure 4 reports the power of the Bonferroni  $t$ -test (Cavanagh, Elliott, and Stock 1995) and the Bonferroni  $Q$ -test described in the last section.<sup>4</sup>

In all cases considered, the Bonferroni  $Q$ -test dominates the Bonferroni  $t$ -test. In fact, the power of the Bonferroni  $Q$ -test comes very close to that of the infeasible  $t$ -test. The power gains of the Bonferroni  $Q$ -test over the Bonferroni  $t$ -test are larger the closer is  $c$  to zero and the larger is  $\delta$  in absolute value. When  $c = -2$  and  $\delta = -0.95$ , the Pitman efficiency is 1.2, which means that the Bonferroni  $t$ -test requires 20% more observations than the Bonferroni  $Q$ -test to achieve 50% power.

In addition to the Bonferroni tests, we also consider the power of Lewellen's (2003) test which is the  $Q$ -test that assumes  $\rho = 1$ . In our notation (24), Lewellen's confidence interval corresponds to  $[\underline{\beta}(1, \alpha_2), \overline{\beta}(1, \alpha_2)]$ . This test can be interpreted as a sup-bound  $Q$ -test, provided that the parameter space is restricted to  $c \in (-\infty, 0]$ , since  $Q(\beta_0, \rho)$  is decreasing in  $\rho$  when  $\delta < 0$ . By construction, the sup-bound  $Q$ -test is the most powerful test when  $c = 0$ . When  $c = -2$  and  $\delta = -0.95$ , the sup-bound  $Q$ -test is undersized when  $b$  is small and has good power when  $b \gg 0$ . When  $c = -2$  and  $\delta = -0.75$ , the power of the sup-bound  $Q$ -test is close to that of the Bonferroni  $Q$ -test. When  $c = -20$ , the sup-bound  $Q$ -test has very poor power. In some sense, the comparison of the sup-bound  $Q$ -test with the Bonferroni tests is unfair because the size of the sup-bound test is greater than 5% when the true autoregressive root is explosive (i.e.  $c > 0$ ), while the Bonferroni tests have the

---

<sup>4</sup>The numerical procedure described in Section 3.3 for the Bonferroni  $Q$ -test is also applied to the Bonferroni  $t$ -test. The significance levels  $\overline{\alpha}_1$  and  $\underline{\alpha}_1$  used in constructing the Dickey-Fuller confidence interval for  $\rho$  are chosen to result in a 5% one-sided test for  $\beta$ , uniformly in  $c \in [-50, 5]$ .

correct size even in the presence of explosive roots.

Against left-sided local alternatives (i.e.  $b < 0$ ), the sup-bound  $t$ -test, which is the  $t$ -test that uses conventional critical values, has correct albeit conservative size. (Recall from Section 3.2 that the left-tailed  $t$ -test is undersized.) Although we do not report the power functions, our computations indicate the Bonferroni tests (based on either the  $t$ -test or the  $Q$ -test) are less undersized than the sup-bound  $t$ -test. Hence, the Bonferroni tests have better power, especially when the predictor variable is persistent (i.e.  $c = -2$ ). The two Bonferroni tests have similar power although the  $t$ -test version has better power when the degree of persistence is low (i.e.  $c = -20$ ).

As revealed by Figure 4, all the feasible tests considered here are biased.<sup>5</sup> That is, the power of the test can be less than the size, for alternatives  $b$  sufficiently close to zero. Recently, there has been progress in the development of unbiased tests for predictive regressions (see Jansson and Moreira (2003) and Polk, Thompson, and Vuolteenaho (2003)). Whether unbiasedness is a desirable property in constructing powerful tests of predictability remains to be seen.

We conclude that the Bonferroni  $Q$ -test has important power advantages over the other feasible tests. Against right-sided alternatives, it has better power than the Bonferroni  $t$ -test, especially when the predictor variable is highly persistent, and it has much better power than the sup-bound  $Q$ -test when the predictor variable is less persistent.

## 4 Predictability of Stock Returns

In this section, we implement our test of predictability in US equity data. We then relate our findings to previous empirical results in the literature.

### 4.1 Data

We use four different series of stock returns, dividend-price ratio, and earnings-price ratio. The first is annual S&P 500 index data (1871–2002) from Global Financial Data since 1926 and Shiller (2000) before then. (The latter is available from the author’s webpage.) The

---

<sup>5</sup>See Lehmann (1986, Chapter 4) for a textbook treatment of unbiasedness for hypothesis testing.

other three series are annual, quarterly, and monthly NYSE/AMEX value-weighted index data (1926–2002) from the Center for Research in Security Prices (CRSP).

Following Campbell and Shiller (1988), the dividend-price ratio is computed as dividends over the past year divided by the current price, and the earnings-price ratio is computed as a moving average of earnings over the past ten years divided by the current price. Since earnings data are not available for the CRSP series, we instead use the corresponding earnings-price ratio from S&P 500. Earnings are available at quarterly frequency since 1935, and annual frequency before then. Shiller (2000) constructs monthly earnings by linear extrapolation. We instead assign quarterly earnings to each month of the quarter since 1935 and annual earnings to each month of the year before then.

To compute excess returns of stocks over a riskfree return, we use the 1-month T-bill rate for the monthly series and the 3-month T-bill rate for the quarterly series. For the annual series, the riskfree return is the return from rolling over the 3-month T-bill every quarter. Since 1926, the T-bill rates are from the CRSP Indices database. For our longer S&P 500 series, we augment this with US Commercial Paper Rates (New York City) from Macaulay (1938), available through NBER’s webpage.

For the three CRSP series, we consider the sub-sample 1952–2002 in addition to the full sample. This allows us to add two additional predictor variables, the 3-month T-bill rate and the long-short yield spread. Following Fama and French (1989), the long yield used in computing the yield spread is Moody’s Seasoned Aaa Corporate Bond Yield. The short rate used is the 1-month T-bill rate. Although data are available before 1952, the nature of the interest rate is very different then due to the Fed’s policy of pegging the interest rate. Following the usual convention, excess returns and the predictor variables are all in logs.

## 4.2 Persistence of Predictor Variables

In Table 3, we report the 95% confidence interval for the autoregressive root  $\rho$  (and the corresponding  $c$ ) for the log dividend-price ratio ( $d - p$ ), the log earnings-price ratio ( $e - p$ ), the 3-month T-bill rate ( $r_3$ ), and the long-short yield spread ( $y - r_1$ ). The confidence interval is computed by the method described in Section 3.2. The autoregressive lag length  $p \in [1, \bar{p}]$  for the predictor variable is estimated by the Bayes Information Criterion (BIC). We set

the maximum lag length  $\bar{p}$  to 4 for annual, 6 for quarterly, and 8 for monthly data. The estimated lag lengths are reported in the fourth column of Table 3.

All of the series are highly persistent, often containing a unit root in the confidence interval. An interesting feature of the confidence intervals for the valuation ratios ( $d - p$  and  $e - p$ ) is that they are sensitive to whether the sample period includes data after 1994. The confidence interval for the sample through 1994 (Panel B) is always less than that for the full sample through 2002 (Panel A). The source of this difference can be explained by Figure 2, which is a time series plot of the valuation ratios at quarterly frequency. Around 1994, these valuation ratios begin to drift down to historical lows, making the processes look more nonstationary. The least persistent series is the yield spread, whose confidence interval never contains a unit root.

The high persistence of these predictor variables suggests that first-order asymptotics, which implies that the  $t$ -statistic is approximately normal in large samples, may be misleading. As discussed in Section 3.2, whether conventional inference based on the  $t$ -test is reliable also depends on the correlation  $\delta$  between the innovations to excess returns and the predictor variable. Hence, we report point estimates of  $\delta$  in the fifth column of Table 3. As expected, the correlations for the valuation ratios are negative and large. This is because movements in stock returns and these valuation ratios mostly come from movements in the stock price. The large magnitude of  $\delta$  suggests that inference based on the conventional  $t$ -test leads to large size distortions. More formally, we fail to reject the null hypothesis that the size distortion is greater than 2.5% using the pretest described in Section 3.2. For the interest rate variables ( $r_3$  and  $y - r_1$ ),  $\delta$  is much smaller. For these predictor variables, the pretest rejects the null hypothesis, which suggests that the conventional  $t$ -test leads to approximately correct inference.

### 4.3 Testing the Predictability of Returns

In this section, we construct valid confidence intervals for  $\beta$  to test the predictability of returns. Based on the power analysis in Section 3.4, our preferred test is the Bonferroni  $Q$ -test.



### 4.3.1 Summary of the Methodology

Our methodology and empirical findings can most easily be explained by the following graphical method, which can be implemented as a series of OLS regressions.

1. Run the regression (12) to obtain the standard error for  $\hat{\beta}$ , denoted as  $\text{SE}(\hat{\beta})$ . Run the regression (17) to obtain the coefficients  $\hat{\psi}_i$  ( $i = 1, \dots, p-1$ ). Using the OLS residuals  $\hat{u}_t$  and  $\hat{e}_t$ , compute  $\hat{\sigma}_u^2 = (T-2)^{-1} \sum_{t=1}^T \hat{u}_t^2$ ,  $\hat{\sigma}_e^2 = (T-2)^{-1} \sum_{t=1}^T \hat{e}_t^2$ ,  $\hat{\sigma}_{ue} = (T-2)^{-1} \sum_{t=1}^T \hat{u}_t \hat{e}_t$ ,  $\hat{\delta} = \hat{\sigma}_{ue} / (\hat{\sigma}_u \hat{\sigma}_e)$ , and  $\hat{\omega}^2 = \hat{\sigma}_e^2 / (1 - \sum_{i=1}^{p-1} \hat{\psi}_i)^2$ .
2. Run the regression (15) to obtain the standard error for  $\hat{\rho}$ , denoted as  $\text{SE}(\hat{\rho})$ . Using the OLS residuals  $\hat{v}_t$ , compute  $\hat{\sigma}_v^2 = (T-2)^{-1} \sum_{t=1}^T \hat{v}_t^2$ .
3. Compute the DF-GLS statistic as follows. Regress  $(x_0, x_1 - \rho_{GLS}x_0, \dots, x_T - \rho_{GLS}x_{T-1})'$  onto  $(1, 1 - \rho_{GLS}, \dots, 1 - \rho_{GLS})'$ , where  $\rho_{GLS} = 1 - 7/T$ , to obtain the coefficient  $\mu_{GLS}$ . Let  $\bar{x}_t = x_t - \mu_{GLS}$ . Run the regression (17) using  $\bar{x}_t$  *without the intercept*. The  $t$ -statistic for  $\theta$  is the DF-GLS statistic.
4. Given the value of the DF-GLS statistic and  $\hat{\delta}$ , use the lookup tables in Campbell and Yogo (2003) to find the appropriate confidence interval  $[\underline{c}, \bar{c}]$  for  $c$ . The confidence interval for  $\rho$  is  $[\underline{\rho}, \bar{\rho}] = [1 + \underline{c}/T, 1 + \bar{c}/T]$ .
5. For each  $\rho \in [\underline{\rho}, \bar{\rho}]$ , compute an equal-tailed 90% confidence interval for  $\beta$  given  $\rho$  as follows. Run the regression (12) with  $r_t - (\hat{\sigma}_{ue} / (\hat{\sigma}_e \hat{\omega}))(x_t - \rho x_{t-1})$  instead of  $r_t$ . Let  $\hat{\beta}(\rho)$  denote the coefficient on  $x_{t-1}$ . The confidence interval for  $\beta$  given  $\rho$  is  $[\underline{\beta}(\rho), \bar{\beta}(\rho)]$ , where

$$\begin{aligned} \underline{\beta}(\rho) &= \hat{\beta}(\rho) + \frac{T-2}{2} \frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e \hat{\omega}} \left( \frac{\hat{\omega}^2}{\hat{\sigma}_v^2} - 1 \right) \text{SE}(\hat{\rho})^2 - z_{0.05} (1 - \hat{\delta}^2)^{1/2} \text{SE}(\hat{\beta}), \\ \bar{\beta}(\rho) &= \hat{\beta}(\rho) + \frac{T-2}{2} \frac{\hat{\sigma}_{ue}}{\hat{\sigma}_e \hat{\omega}} \left( \frac{\hat{\omega}^2}{\hat{\sigma}_v^2} - 1 \right) \text{SE}(\hat{\rho})^2 + z_{0.05} (1 - \hat{\delta}^2)^{1/2} \text{SE}(\hat{\beta}). \end{aligned}$$

6. Plot  $[\underline{\beta}(\rho), \bar{\beta}(\rho)]$  against  $\rho$  for all  $\rho \in [\underline{\rho}, \bar{\rho}]$ .

In step 1, the autoregressive lag length  $p$  can be estimated consistently by BIC. When the predictor variable is an AR(1) (i.e.  $p = 1$ ), regressions (15) and (17) are equivalent, so step

2 can be eliminated. In addition, the formulas in step 5 simplify since  $\hat{\omega}^2 = \hat{\sigma}_v^2 = \hat{\sigma}_e^2$  in that case. In practice, we only need to compute the confidence interval  $[\underline{\beta}(\rho), \bar{\beta}(\rho)]$  at the end points of  $[\underline{\rho}, \bar{\rho}]$  since  $\underline{\beta}(\rho)$  and  $\bar{\beta}(\rho)$  are linear in  $\rho$ . The 90% Bonferroni confidence interval  $[\underline{\beta}(\bar{\rho}), \bar{\beta}(\underline{\rho})]$  corresponds to a 10% two-sided test or a 5% one-sided test for predictability.

In reporting our confidence interval for  $\beta$ , we will scale it by  $\hat{\sigma}_e/\hat{\sigma}_u$ . In other words, we report the confidence interval for  $\tilde{\beta} = (\sigma_e/\sigma_u)\beta$  instead of  $\beta$ . Although this normalization does not affect inference, it is a more natural way to report the empirical results for two reasons. First,  $\tilde{\beta}$  has a natural interpretation as the coefficient in (12) when the innovations in (12) and (17) are normalized to have unit variance. This is in the spirit of the simple regression model (1)–(2), which assumed unit variance in the innovations. Second, by the equality

$$\tilde{\beta} = \frac{\sigma(\mathbf{E}_{t-1}r_t - \mathbf{E}_{t-2}r_t)}{\sigma(r_t - \mathbf{E}_{t-1}r_t)},$$

$\tilde{\beta}$  can be interpreted as the standard deviation of the change in expected returns relative to the standard deviation of the innovation to returns. To simplify notation, we will use  $\beta$  to denote  $\tilde{\beta}$  throughout the rest of the paper.

#### 4.3.2 Empirical Findings

In Figure 5, we plot the Bonferroni confidence interval, using the annual and quarterly CRSP series (1926–2002), when the predictor variable is the dividend-price ratio or the earnings-price ratio. The thick lines represent the confidence interval based on the Bonferroni  $Q$ -test, and the thin lines represent the confidence interval based on the Bonferroni  $t$ -test. Because of the asymmetry in the null distribution of the  $t$ -statistic, the confidence interval for  $\rho$  used for the right-tailed Bonferroni  $t$ -test differs from that used for the left-tailed test. This explains why the length of the lower bound of the interval, corresponding to the right-tailed test, can differ from the upper bound, corresponding to the left-tailed test. The application of the Bonferroni  $Q$ -test is new, but the Bonferroni  $t$ -test has been applied previously by Torous, Valkanov, and Yan (2001). We report the latter for the purpose of comparison.

For the annual dividend-price ratio in Panel A, the Bonferroni confidence interval for  $\beta$  based on the  $Q$ -test lies strictly above zero. Hence, we can reject the null  $\beta = 0$  against the alternative  $\beta > 0$  at the 5% level. The Bonferroni confidence interval based on the  $t$ -test,

however, includes  $\beta = 0$ . Hence, we cannot reject the null of no predictability using the Bonferroni  $t$ -test. This can be interpreted in light of the power comparisons in Figure 4. From Table 3,  $\hat{\delta} = -0.721$  and the confidence interval for  $c$  is  $[-7.343, 3.781]$ . In this region of the parameter space, the Bonferroni  $Q$ -test is more powerful than the Bonferroni  $t$ -test against right-sided alternatives, resulting in a tighter confidence interval.

For the quarterly dividend-price ratio in Panel C, the evidence for predictability is weaker. In the relevant range of the confidence interval for  $\rho$ , the confidence interval for  $\beta$  contains zero for both the Bonferroni  $Q$ -test and  $t$ -test, although the confidence interval is again tighter for the  $Q$ -test. Using the Bonferroni  $Q$ -test, the confidence interval for  $\beta$  lies above zero when  $\rho \leq 0.988$ . This means that if the true  $\rho$  is less than 0.988, we can reject the null hypothesis  $\beta = 0$  against the alternative  $\beta > 0$  at the 5% level. On the other hand, if  $\rho > 0.988$ , the confidence interval includes  $\beta = 0$ , so we cannot reject the null. Since there is uncertainty over the true value of  $\rho$ , we cannot reject the null of no predictability.

In Panel B, we test for predictability in annual data using the earnings-price ratio as the predictor variable. We find that stock returns are predictable with the Bonferroni  $Q$ -test, but not with the Bonferroni  $t$ -test. In Panel D, we obtain the same results at the quarterly frequency. Again, the Bonferroni  $Q$ -test gives tighter confidence intervals due to better power, which is empirically relevant for detecting predictability.

In Figure 6, we repeat the same exercise as Figure 5, using the quarterly CRSP series in the sub-sample 1952–2002. We report the plots for all four of our predictor variables: A) the dividend-price ratio, B) the earnings-price ratio, C) the T-bill rate, and D) the yield spread.

For the dividend-price ratio, we find evidence for predictability if  $\rho \leq 1.004$ . This means that if we are willing to rule out explosive roots, confining attention to the area of the figure to the left of the vertical line at  $\rho = 1$ , we can conclude that returns are predictable with the dividend-price ratio. The confidence interval for  $\rho$ , however, includes explosive roots, so we cannot impose  $\rho \leq 1$  without using prior information about the behavior of the dividend-price ratio.

The earnings-price ratio is a less successful predictor variable in this sub-sample. We find that  $\rho$  must be less than 0.997 before we can conclude that the earnings-price ratio predicts returns. Taking account of the uncertainty in the true value of  $\rho$ , we cannot reject

the null hypothesis  $\beta = 0$ . The weaker evidence for predictability in the period since 1952 is partly due to the fact that the valuation ratios appear more persistent when restricted to this sub-sample. The confidence intervals therefore contain rather large values of  $\rho$  that were excluded in Figure 5.

For the T-bill rate, the Bonferroni confidence interval for  $\beta$  lies strictly below zero for both the  $Q$ -test and the  $t$ -test over the entire confidence interval for  $\rho$ . For the yield spread, the evidence for predictability is similarly strong, with the confidence interval strictly above zero over the entire range of  $\rho$ . The power advantage of the Bonferroni  $Q$ -test over the Bonferroni  $t$ -test is small when  $\delta$  is small in absolute value, so these tests result in very similar confidence intervals.

In Table 4, we report the complete set of results in tabular form. In the fifth column of the table, we report the 90% Bonferroni confidence intervals for  $\beta$  using the  $t$ -test. In the sixth column, we report the 90% Bonferroni confidence interval using the  $Q$ -test. In terms of Figures 5–6, we simply report the minimum and maximum values of  $\beta$  for each of the tests.

Focusing first on the full-sample results in Panel A, the Bonferroni  $Q$ -test rejects the null of no predictability for the earnings-price ratio ( $e - p$ ) at all frequencies. For the dividend-price ratio ( $d - p$ ), we fail to reject the null except for the annual CRSP series. Using the Bonferroni  $t$ -test, we always fail to reject the null due to its poor power relative to the Bonferroni  $Q$ -test.

In the sub-sample through 1994, reported in Panel B, the results are qualitatively similar. In particular, the Bonferroni  $Q$ -test finds predictability with the earnings-price ratio at all frequencies. Interestingly, the Bonferroni  $t$ -test also finds predictability in this sub-sample, although the confidence intervals are still wider than those of the Bonferroni  $Q$ -test. In this sub-sample, the evidence for predictability is sufficiently strong that a relatively inefficient test can also find predictability.

In Panel C, we report the results for the sub-sample since 1952. In this sub-sample, we cannot reject the null hypothesis for the valuation ratios ( $d - p$  and  $e - p$ ). For the T-bill rate and the yield spread ( $r_3$  and  $y - r_1$ ), however, we reject the null hypothesis except at annual frequency.

As we have seen in Figure 6, the weak evidence for predictability in this sub-sample

arises from the fact that the confidence intervals for  $\rho$  contain explosive roots. If we could obtain tighter confidence intervals for  $\rho$  that exclude these values, the lower bound of the confidence intervals for  $\beta$  would rise, strengthening the evidence for predictability. In the last column of Table 4, we report the lower bound of the confidence interval for  $\beta$  at  $\rho = 1$ . This corresponds to Lewellen’s (2003) sup-bound  $Q$ -test, which restricts the parameter space to  $\rho \leq 1$ . In terms of Figures 5–6, this is equivalent to discarding the region of the plots where  $\rho > 1$ . Under this restriction, the lower bound of the confidence interval for the dividend-price ratio lies above zero at all frequencies. The dividend-price ratio therefore predicts returns in the sub-sample since 1952 provided that its autoregressive root is not explosive, consistent with Lewellen’s findings.

To summarize the empirical results, we find reliable evidence for predictability with the earnings-price ratio, the T-bill rate, and the yield spread. The evidence for predictability with the dividend-price ratio is weaker, and we do not find unambiguous evidence for predictability using our Bonferroni  $Q$ -test. The Bonferroni  $Q$ -test gives tighter confidence intervals than the Bonferroni  $t$ -test due to better power. The power gain is empirically important in the full sample through 2002.

## 4.4 Connection to Previous Empirical Findings

The empirical literature on the predictability of returns is rather large, and in this section, we attempt to interpret the main findings in light of our analysis in the last section.

### 4.4.1 $t$ -test

The earliest and the most intuitive approach to testing predictability is to run the predictive regression and to compute the  $t$ -statistic. One would then reject the null hypothesis  $\beta = 0$  against the alternative  $\beta > 0$  at the 5% level if the  $t$ -statistic is greater than 1.645. In the third column of Table 4, we report the  $t$ -statistics from the predictive regressions. Using the conventional critical value, the  $t$ -statistics are mostly “significant”, often greater than two and sometimes greater than three. Comparing the full sample through 2002 (Panel A) and the sub-sample through 1994 (Panel B), the evidence for predictability appears to

have weakened in the last eight years. In the late 1990's, stock returns were high when the valuation ratios were at historical lows. Hence, the evidence for predictability “went in the wrong direction”.

However, one may worry about statistical inference that is so sensitive to an addition of 8 observations to a sample of 115 (for S&P 500) or an addition of 32 data points to a sample of 273 (for quarterly CRSP). In fact, this sensitivity is evidence for the failure of first-order asymptotics. Intuitively, when a predictor variable is persistent, its sample moments can change dramatically with an addition of a few data points. Since the  $t$ -statistic measures the covariance of excess returns with the lagged predictor variable, its value is sensitive to persistent deviations in the predictor variable from the mean. This is what happened in the late 1990's when valuation ratios reached historical lows. Tests that are derived from local-to-unity asymptotics take this persistence into account and hence lead to correct inference.

Using the Bonferroni  $Q$ -test, which is robust to the persistence problem, we find that the earnings-price ratio predicts returns in both the full sample and the sub-sample through 1994. There appears to be some empirical content in the claim that the evidence for predictability has weakened, with the Bonferroni confidence interval based on the  $Q$ -test shifting toward zero. Using the Bonferroni confidence interval based on the  $t$ -test, we reject the null of no predictability in the sub-sample through 1994, but not in the full sample. The “weakened” evidence for predictability in the recent years puts a premium on the efficiency of test procedures.

As additional evidence for the failure of first-order asymptotics, we report the OLS point estimates of  $\beta$  in the fourth column of Table 4. As equations (22)–(23) show, the point estimate  $\hat{\beta}$  does not necessarily lie in the center of the robust confidence interval for  $\beta$ . Indeed,  $\hat{\beta}$  for the valuation ratios are usually closer to the upper bound of the Bonferroni confidence interval based on the  $Q$ -test, and in a few cases (dividend-price ratio in Panel C),  $\hat{\beta}$  falls strictly above the confidence interval. This is a consequence of the upward finite-sample bias of the OLS estimator arising from the persistence of these predictor variables (see Stambaugh (1999) and Lewellen (2003)).

One way to interpret the  $t$ -test based on the conventional critical value (1.645 for a 5% one-sided test) is the Bayesian interpretation. Suppose  $\delta = -0.9$ , which is a reasonable

value for the valuation ratios. As reported in Table 1, the unknown persistence parameter  $c$  must be less than  $-70$  for the size distortion of the  $t$ -test to be less than 2.5%. Hence, if a researcher has prior information that  $c < -70$ , he can proceed with the  $t$ -test using the critical value 1.645. Our empirical findings in Figures 5–6 confirm that there is strong evidence for predictability with the valuation ratios when  $\rho \ll 1$ . The problem with such inference is that the lower bound of the confidence interval for  $c$  is much greater than  $-70$ , so it is hard to reconcile the prior belief in a low  $c$  with the observed persistence of the valuation ratios.

For the interest rate variables, the correlation  $\delta$  is sufficiently small that conventional inference based on the  $t$ -test leads to approximately correct inference. This is confirmed in Panel C of Table 4, where inference based on the conventional  $t$ -test agrees with that based on the Bonferroni  $Q$ -test.

#### 4.4.2 Long-Horizon Tests

Some authors, notably Fama and French (1988) and Campbell and Shiller (1988), have explored the behavior of stock returns at lower frequencies by regressing long-horizon returns onto financial variables. In annual data, the dividend-price ratio has a smaller autoregressive root than it does in monthly data and is less persistent in that sense. Over several years, the ratio has an even smaller autoregressive root. Unfortunately, this does not eliminate the statistical problem caused by persistence because the effective sample size shrinks as one increases the horizon of the regression.

Recently, a number of authors have pointed out that the finite-sample distribution of the long-horizon regression coefficient and its associated  $t$ -statistic can be quite different from the asymptotic distribution due to persistence in the regressor and overlap in the returns data. (See Hodrick (1992), Nelson and Kim (1993), Ang and Bekaert (2001) for computational results and Valkanov (2003) and Torous, Valkanov, and Yan (2001) for analytical results.) Accounting for these problems, Torous, Valkanov, and Yan (2001) find no evidence for predictability at long horizons using many of the popular predictor variables. In fact, they find no evidence for predictability at any horizon or time period, except at quarterly and annual frequency in 1952–1994.

Long-horizon regressions can also be understood as a way to reduce the noise in stock returns, because under the alternative hypothesis that returns are predictable, the variance of the return increases less than proportionally with the investment horizon (see Campbell, Lo, and MacKinlay (1997, Chapter 7) and Campbell (2001)). The procedures developed in this paper and in Lewellen (2003) have the important advantage that they reduce noise not only under the alternative, but also under the null. Thus they increase power against local alternatives, while long-horizon regression tests do not.

#### 4.4.3 Other Tests

In this section, we discuss three recent papers that have taken the issue of persistence seriously to develop tests that have the correct size even if the predictor variable is highly persistent or  $I(1)$ .

Lewellen (2003) proposes to test the predictability of returns by computing the  $Q$ -statistic evaluated at  $\beta_0 = 0$  and  $\rho = 1$  (i.e.  $Q(0, 1)$ ). His test procedure rejects  $\beta = 0$  against the one-sided alternative  $\beta > 0$  at the  $\alpha$ -level if  $Q(0, 1) > z_\alpha$ . Since the null distribution of  $Q(0, 1)$  is standard normal under local-to-unity asymptotics, Lewellen's test procedure has correct size as long as  $\rho = 1$ . If  $\rho \neq 1$ , this procedure does not in general have the correct size. However, Lewellen's procedure is a valid (although conservative) one-sided test as long as  $\delta \leq 0$  and  $\rho \leq 1$ . As we have shown in Panel C of Table 4, the 5% one-sided test using the monthly dividend-price ratio rejects when  $\rho = 1$ , confirming Lewellen's empirical findings.

Based on finance theory, it is reasonable to assume that the dividend-price ratio is mean reverting, at least in the very long run. However, we may not necessarily want to impose Lewellen's parametric assumption that the dividend-price ratio is an  $AR(1)$  with  $\rho \leq 1$ . In the absence of knowledge of the true data-generating process, the purpose of the parametric model (15)–(16) is to provide a flexible framework to approximate the dynamics of the predictor variable in finite samples. Allowing for the possibility that  $\rho > 1$  can be an important part of that flexibility, especially in light of the recent behavior of the dividend-price ratio. In addition, we allow for possible short-run dynamics in the predictor variable by considering an  $AR(p)$ , which Lewellen rules out by imposing a strict  $AR(1)$ .

Another issue that arises with Lewellen's test is that of power. As shown in Figure 4,



the test can have poor power when the predictor variable is stationary (i.e.  $\rho < 1$ ). For instance, the annual earnings-price ratio for the S&P 500 index has 95% confidence interval  $[0.768, 0.965]$  for  $\rho$ . As reported in Panel A of Table 4, the lower bound of the confidence interval for  $\beta$  using the Bonferroni  $Q$ -test is 0.043, rejecting the null of no predictability. However, the  $Q$ -test at  $\rho = 1$  results in a lower bound of -0.023, failing to reject the null. Therefore, the poor power of Lewellen’s procedure leads to the false conclusion that the earning-price ratio does not predict returns at annual frequency. Similarly, Lewellen’s procedure always leads to wider confidence intervals than the Bonferroni  $Q$ -test in the sub-sample through 1994, when the valuation ratios are less persistent.

Torous, Valkanov, and Yan (2001) develop a test of predictability that is conceptually similar to ours, constructing Bonferroni confidence intervals for  $\beta$ . One difference from our approach is that they construct the confidence interval for  $\rho$  using the augmented Dickey-Fuller test, rather than the more powerful DF-GLS test of Elliott, Rothenberg, and Stock (1996). The second difference is that they use the long-horizon  $t$ -test, instead of the more powerful  $Q$ -test, for constructing the confidence interval of  $\beta$  given  $\rho$ . Their choice of the long-horizon  $t$ -test is motivated by their objective of highlighting the pitfalls of long-horizon regressions.

A key insight in Torous, Valkanov, and Yan (2001) is that the evidence for the predictability of returns depends critically on the unknown degree of persistence of the predictor variable. Because we cannot estimate the degree of persistence consistently, the evidence for predictability can be ambiguous. This point is illustrated in Figures 5–6, where we find that the dividend-price ratio predicts returns if its autoregressive root  $\rho$  is sufficiently small. In this paper, we have confirmed their finding that the evidence for predictability by the dividend-price ratio is weak once its persistence has been properly accounted for.

A different approach to dealing with the problem of persistence is to ignore the data on predictor variables and to base inference solely on the returns data. Under the null that returns are not predictable by a persistent predictor variable, returns should behave like a stationary process. Under the alternative of predictability, returns should have a unit or a near-unit root. Using this approach, Lanne (2002) fails to reject the null of no predictability. However, his test is conservative in the sense that it has poor power when the predictor

variable is persistent but not close enough to being integrated.<sup>6</sup> Lanne’s empirical finding agrees with ours and those of Torous, Valkanov, and Yan (2001). From Figures 5–6, we see that the valuation ratios predict returns provided that their degree of persistence is sufficiently small. In addition, we find evidence for predictability with the yield spread, which has a relatively small degree of persistence compared to the valuation ratios. Lanne’s test would fail to detect predictability by less persistent variables like the yield spread.

## 5 Conclusion

The hypothesis that stock returns are predictable at long horizons has been called a “new fact in finance” (Cochrane 1999). That the predictability of stock returns is now widely accepted by financial economists is remarkable given the long tradition of the “random walk” model of stock prices. In this paper, we have shown that there is indeed evidence for predictability, but it is more challenging to detect than previous studies may have suggested. Most popular and economically sensible candidates for predictor variables (such as the dividend-price ratio, earnings-price ratio, or measures of the interest rate) are highly persistent. When the predictor variable is persistent, the distribution of the  $t$ -statistic is nonstandard, which can lead to over-rejection of the null hypothesis using conventional critical values.

In this paper, we have developed a pretest to determine when the conventional  $t$ -test leads to misleading inferences. Using the pretest, we find that the  $t$ -test leads to correct inference for the short-term interest rate and the long-short yield spread. Persistence is not a problem for these interest rate variables because their innovations have sufficiently low correlation with the innovations to stock returns. Using the  $t$ -test with conventional critical values, we find that these interest rate variables predict returns in the post-1952 sample.

For the dividend-price ratio and the smoothed earnings-price ratio, persistence is an issue since their innovations are highly correlated with the innovations to stock returns. Using our pretest, we find that the conventional  $t$ -test can lead to misleading inferences for these

---

<sup>6</sup>In fact, Campbell, Lo, and MacKinlay (1997, Chapter 7) construct an example in which returns are univariate white noise but are predictable using a stationary variable with an arbitrary autoregressive coefficient.

valuation ratios. In this paper, we have developed an efficient test of predictability that leads to correct inference regardless of the degree of persistence of the predictor variable. Over the full sample, our test reveals that the earnings-price ratio reliably predicts returns at various frequencies (annual to monthly), while the dividend-price ratio predicts returns only at annual frequency. In the post-1952 sample, the evidence for predictability is weaker, but the dividend-price ratio predicts returns if we can rule out explosive autoregressive roots.

Taken together, these results suggest that there is a predictable component in stock returns, but one that is difficult to detect without careful use of efficient statistical tests.

## Appendix

In this appendix, we derive the asymptotic distribution of the  $t$ -statistic and the  $Q$ -statistic under the local alternative  $\beta = \beta_0 + b/T$ . These asymptotic representations are used to compute the power functions of the various test procedures in Section 3.4.

The  $t$ -statistic can be written as

$$t(\beta_0) = \frac{b(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\sigma_u} + \frac{T^{-1} \sum_{t=1}^T x_{t-1}^{\mu} u_t}{\sigma_u (T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}.$$

By Lemma 1 (see also Cavanagh, Elliott, and Stock (1995)),

$$t(\beta_0) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u} + \delta \frac{\tau_c}{\kappa_c} + (1 - \delta^2)^{1/2} Z, \quad (27)$$

where  $Z$  is a standard normal random variable independent of  $(W_e(s), J_c(s))$ .

Let  $\tilde{\rho} = 1 + \tilde{c}/T$  and define the statistic

$$Q(\beta_0, \tilde{\rho}) = \frac{\sum_{t=1}^T x_{t-1}^{\mu} [r_t - \beta_0 x_{t-1} - \frac{\sigma_{ue}}{\sigma_e \omega} (x_t - \tilde{\rho} x_{t-1})] + \frac{T}{2} \frac{\sigma_{ue}}{\sigma_e \omega} (\omega^2 - \sigma_v^2)}{\sigma_u (1 - \delta^2)^{1/2} (\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}. \quad (28)$$

The three types of  $Q$ -test considered in Section 3.4 correspond to special cases of this test statistic.

1. Infeasible  $Q$ -test:  $\tilde{\rho} = \rho = 1 + c/T$ .
2. Bonferroni  $Q$ -test:  $\tilde{\rho} = \bar{\rho} = 1 + \bar{c}/T$ , where  $\bar{c}$  depends on the DF-GLS statistic and  $\delta$ .
3. Sup-bound  $Q$ -test:  $\tilde{\rho} = 1$ .

The statistic (28) can be written as

$$\begin{aligned} Q(\beta_0, \tilde{\rho}) &= \frac{b(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\sigma_u (1 - \delta^2)^{1/2}} + \frac{\delta(\tilde{c} - c)(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\omega (1 - \delta^2)^{1/2}} \\ &\quad + \frac{T^{-1} \sum_{t=1}^T x_{t-1}^{\mu} (u_t - \frac{\sigma_{ue}}{\sigma_e \omega} v_t) + \frac{1}{2} \frac{\sigma_{ue}}{\sigma_e \omega} (\omega^2 - \sigma_v^2)}{\sigma_u (1 - \delta^2)^{1/2} (T^{-2} \sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}. \end{aligned}$$

By Lemma 1,

$$Q(\beta_0, \rho) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u (1 - \delta^2)^{1/2}} + \frac{\delta(\tilde{c} - c)\kappa_c}{(1 - \delta^2)^{1/2}} + Z. \quad (29)$$

The power function for the right-tailed test (i.e.  $b > 0$ ) is therefore given by

$$\pi_Q(b) = \mathbf{E} \left[ \Phi \left( z_\alpha - \frac{b\omega\kappa_c}{\sigma_u (1 - \delta^2)^{1/2}} - \frac{\delta(\tilde{c} - c)\kappa_c}{(1 - \delta^2)^{1/2}} \right) \right], \quad (30)$$

where the expectation is taken over the distribution of  $(W_e(s), J_c(s))$ .

Following Stock (1991, Appendix B), the limiting distributions (27) and (29) are approximated by Monte Carlo simulation. We generate 20,000 realizations of the Gaussian AR(1) (i.e. model (2)) with  $T = 500$ ,  $\rho = 1 + c/T$ , and  $x_0 = 0$ . The distribution of  $\kappa_c$  is approximated by  $(T^{-2} \sum_{t=1}^T x_{t-1}^{\mu^2})^{1/2}$ , and  $\tau_c$  is approximated by  $T^{-1} \sum_{t=1}^T x_{t-1}^{\mu} e_t$ .

## References

- Ang, Andrew, and Geert Bekaert, 2001, Stock return predictability: Is it there?, Working paper, Columbia Business School.
- Billingsley, Patrick, 1999, *Convergence of Probability Measures*. Wiley Series in Probability and Statistics (John Wiley & Sons, Inc.: New York) second edn.
- Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.
- , 2001, Why long horizons? A study of power against persistent alternatives, *Journal of Empirical Finance* 8, 459–491.
- , Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets* (Princeton University Press: Princeton, NJ).
- Campbell, John Y., and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *The Journal of Finance* 43, 661–676.
- Campbell, John Y., and Motohiro Yogo, 2003, Appendix to “Efficient tests of stock return predictability”, Working paper, Harvard University.
- Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.
- Cochrane, John H., 1999, New facts in finance, NBER Working Paper 7169.
- Cox, D. R., and D. V. Hinkley, 1974, *Theoretical Statistics* (Chapman and Hall: London).
- Dickey, David A., and Wayne A. Fuller, 1981, Likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica* 49, 1057–1072.
- Elliott, Graham, Thomas J. Rothenberg, and James H. Stock, 1996, Efficient tests for an autoregressive unit root, *Econometrica* 64, 813–836.
- Elliott, Graham, and James H. Stock, 1994, Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory* 10, 672–700.

- , 2001, Confidence intervals for autoregressive coefficients near one, *Journal of Econometrics* 103, 155–181.
- Evans, G. B. A., and N. E. Savin, 1981, Testing for unit roots: 1, *Econometrica* 49, 753–779.
- , 1984, Testing for unit roots: 2, *Econometrica* 52, 1241–1270.
- Fama, Eugene F., and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–24.
- , 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.
- Hodrick, Robert J., 1992, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, *The Review of Financial Studies* 5, 357–386.
- Jansson, Michael, and Marcelo J. Moreira, 2003, Optimal inference in regression models with nearly integrated regressors, Working paper, Harvard University.
- Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.
- Lanne, Markku, 2002, Testing the predictability of stock returns, *The Review of Economics and Statistics* 84, 407–415.
- Lehmann, E. L., 1986, *Testing Statistical Hypotheses*. Springer Texts in Statistics (Springer-Verlag: New York) second edn.
- , 1999, *Elements of Large Sample Theory*. Springer Texts in Statistics (Springer-Verlag: New York) second edn.
- Lewellen, Jonathan, 2003, Predicting returns with financial ratios, *Journal of Financial Economics*, forthcoming.
- Macaulay, F. R., 1938, *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856* (National Bureau of Economic Research: New York).

- Mankiw, N. Gregory, and Matthew D. Shapiro, 1986, Do we reject too often? Small sample properties of tests of rational expectations models, *Economics Letters* 20, 139–145.
- Nelson, Charles R., and Myung J. Kim, 1993, Predictable stock returns: The role of small sample bias, *The Journal of Finance* 48, 641–661.
- Phillips, Peter C. B., 1987, Towards a unified asymptotic theory for autoregression, *Biometrika* 74, 535–547.
- , and Pierre Perron, 1988, Testing for a unit root in time series regression, *Biometrika* 75, 335–346.
- Polk, Christopher, Samuel Thompson, and Tuomo Vuolteenaho, 2003, New forecasts of the equity premium, Working paper, Harvard University.
- Richardson, Matthew, and James H. Stock, 1989, Drawing inferences from statistics based on multiyear asset returns, *Journal of Financial Economics* 25, 323–348.
- Shiller, Robert J., 2000, *Irrational Exuberance* (Princeton University Press: Princeton, NJ).
- Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375–421.
- Stock, James H., 1991, Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, *Journal of Monetary Economics* 28, 435–459.
- , 1994, Unit roots, structural breaks and trends, in Robert F. Engle, and Daniel L. McFadden, ed.: *Handbook of Econometrics*, vol. 4 . chap. 46, pp. 2739–2841 (Elsevier Science Publishers: New York).
- Torous, Walter, Rossen Valkanov, and Shu Yan, 2001, On predicting stock returns with nearly integrated explanatory variables, *The Journal of Business*, forthcoming.
- Valkanov, Rossen, 2003, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics* 68, 201–232.



Table 1: Parameters Leading to Size Distortion of the One-Sided  $t$ -test

This table reports the regions of the parameter space where the actual size of the nominal 5%  $t$ -test is greater than 7.5%. The null hypothesis is  $\beta = \beta_0$  against the alternative  $\beta > \beta_0$ . For a given  $\delta$ , the size of the  $t$ -test is greater than 7.5% if  $c \in (c_{\min}, c_{\max})$ . Size is less than 7.5% for all  $c$  if  $\delta \leq -0.125$ .

$\delta$	$c_{\min}$	$c_{\max}$	$\delta$	$c_{\min}$	$c_{\max}$
-1.000	-83.088	8.537	-0.550	-28.527	6.301
-0.975	-81.259	8.516	-0.525	-27.255	6.175
-0.950	-79.318	8.326	-0.500	-25.942	6.028
-0.925	-76.404	8.173	-0.475	-23.013	5.868
-0.900	-69.788	7.977	-0.450	-19.515	5.646
-0.875	-68.460	7.930	-0.425	-17.701	5.435
-0.850	-63.277	7.856	-0.400	-14.809	5.277
-0.825	-59.563	7.766	-0.375	-13.436	5.111
-0.800	-58.806	7.683	-0.350	-11.884	4.898
-0.775	-57.618	7.585	-0.325	-10.457	4.682
-0.750	-51.399	7.514	-0.300	-8.630	4.412
-0.725	-50.764	7.406	-0.275	-6.824	4.184
-0.700	-42.267	7.131	-0.250	-5.395	3.934
-0.675	-41.515	6.929	-0.225	-4.431	3.656
-0.650	-40.720	6.820	-0.200	-3.248	3.306
-0.625	-36.148	6.697	-0.175	-1.952	2.800
-0.600	-33.899	6.557	-0.150	-0.614	2.136
-0.575	-31.478	6.419	-0.125	—	—

Table 2: Significance Level of the DF-GLS Confidence Interval for the Bonferroni  $Q$ -test

This table reports the significance level of the confidence interval for the largest autoregressive root  $\rho$ , computed by inverting the DF-GLS test, which sets the size of the one-sided Bonferroni  $Q$ -test to 5%. Using the notation (24), the confidence interval  $C_\rho(\alpha_1) = [\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)]$  for  $\rho$  results in a 90% Bonferroni confidence interval  $C_\beta(0.1)$  for  $\beta$  when  $\alpha_2 = 0.1$ .

$\delta$	$\underline{\alpha}_1$	$\overline{\alpha}_1$	$\delta$	$\underline{\alpha}_1$	$\overline{\alpha}_1$
-0.999	0.050	0.055	-0.500	0.080	0.280
-0.975	0.055	0.080	-0.475	0.085	0.285
-0.950	0.055	0.100	-0.450	0.085	0.295
-0.925	0.055	0.115	-0.425	0.090	0.310
-0.900	0.060	0.130	-0.400	0.090	0.320
-0.875	0.060	0.140	-0.375	0.095	0.330
-0.850	0.060	0.150	-0.350	0.100	0.345
-0.825	0.060	0.160	-0.325	0.100	0.355
-0.800	0.065	0.170	-0.300	0.105	0.360
-0.775	0.065	0.180	-0.275	0.110	0.370
-0.750	0.065	0.190	-0.250	0.115	0.375
-0.725	0.065	0.195	-0.225	0.125	0.380
-0.700	0.070	0.205	-0.200	0.130	0.390
-0.675	0.070	0.215	-0.175	0.140	0.395
-0.650	0.070	0.225	-0.150	0.150	0.400
-0.625	0.075	0.230	-0.125	0.160	0.405
-0.600	0.075	0.240	-0.100	0.175	0.415
-0.575	0.075	0.250	-0.075	0.190	0.420
-0.550	0.080	0.260	-0.050	0.215	0.425
-0.525	0.080	0.270	-0.025	0.250	0.435

Table 3: Estimates of the Model Parameters

This table reports estimates of the parameters for the predictive regression model. Returns are for annual S&P 500 and annual, quarterly, and monthly CRSP value-weighted index. The predictor variables are the log dividend-price ratio ( $d - p$ ), the log earnings-price ratio ( $e - p$ ), the 3-month T-bill rate ( $r_3$ ), and the long-short yield spread ( $y - r_1$ ).  $p$  is the estimated autoregressive lag length for the predictor variable, and  $\delta$  is the estimated correlation between the innovations to returns and the predictor variable. The last two columns are the 95% confidence intervals for the largest autoregressive root ( $\rho$ ) and the corresponding local-to-unity parameter ( $c$ ) for each of the predictor variables, computed using the DF-GLS statistic.

Series	Sample (Obs)	Variable	$p$	$\delta$	DF-GLS	95% CI: $\rho$	95% CI: $c$
Panel A: Full Sample							
S&P 500	1880–2002	$d - p$	3	-0.846	-0.855	[0.949,1.033]	[-6.107,4.020]
	(123)	$e - p$	1	-0.962	-2.888	[0.768,0.965]	[-28.262,-4.232]
Annual	1926–2002	$d - p$	1	-0.721	-1.033	[0.903,1.050]	[-7.343,3.781]
	(77)	$e - p$	1	-0.957	-2.229	[0.748,1.000]	[-19.132,-0.027]
Quarterly	1926–2002	$d - p$	1	-0.942	-1.696	[0.957,1.007]	[-13.081,2.218]
	(305)	$e - p$	1	-0.986	-2.191	[0.939,1.000]	[-18.670,0.145]
Monthly	1926–2002	$d - p$	2	-0.950	-1.657	[0.986,1.003]	[-12.683,2.377]
	(913)	$e - p$	1	-0.987	-1.859	[0.984,1.002]	[-14.797,1.711]
(continued on the next page)							

Series	Sample (Obs)	Variable	$p$	$\delta$	DF-GLS	95% CI: $\rho$	95% CI: $c$
Panel B: Sample through 1994							
S&P 500	1880–1994	$d - p$	3	-0.836	-2.002	[0.854,1.010]	[-16.391,1.079]
	(115)	$e - p$	1	-0.958	-3.519	[0.663,0.914]	[-38.471,-9.789]
Annual	1926–1994	$d - p$	1	-0.693	-2.081	[0.745,1.010]	[-17.341,0.690]
	(69)	$e - p$	1	-0.959	-2.859	[0.591,0.940]	[-27.808,-4.074]
Quarterly	1926–1994	$d - p$	1	-0.941	-2.635	[0.910,0.991]	[-24.579,-2.470]
	(273)	$e - p$	1	-0.988	-2.827	[0.900,0.986]	[-27.322,-3.844]
Monthly	1926–1994	$d - p$	2	-0.948	-2.551	[0.971,0.998]	[-23.419,-1.914]
	(817)	$e - p$	2	-0.983	-2.600	[0.970,0.997]	[-24.105,-2.240]
Panel C: Sample from 1952							
Annual	1952–2002	$d - p$	1	-0.749	-0.462	[0.917,1.087]	[-4.131,4.339]
	(51)	$e - p$	1	-0.955	-1.522	[0.773,1.056]	[-11.354,2.811]
		$r_3$	1	0.006	-1.762	[0.725,1.040]	[-13.756,1.984]
		$y - r_1$	1	-0.243	-3.121	[0.363,0.878]	[-31.870,-6.100]
Quarterly	1952–2002	$d - p$	1	-0.977	-0.392	[0.981,1.022]	[-3.844,4.381]
	(204)	$e - p$	1	-0.980	-1.195	[0.958,1.017]	[-8.478,3.539]
		$r_3$	4	-0.095	-1.572	[0.941,1.013]	[-11.825,2.669]
		$y - r_1$	2	-0.100	-2.765	[0.869,0.983]	[-26.375,-3.347]
Monthly	1952–2002	$d - p$	1	-0.967	-0.275	[0.994,1.007]	[-3.365,4.451]
	(612)	$e - p$	1	-0.982	-0.978	[0.989,1.006]	[-6.950,3.857]
		$r_3$	2	-0.071	-1.569	[0.981,1.004]	[-11.801,2.676]
		$y - r_1$	1	-0.066	-4.368	[0.911,0.968]	[-54.471,-19.335]

Table 4: Tests of Predictability

This table reports statistics used to infer the predictability of returns. Returns are for annual S&P 500 and annual, quarterly, and monthly CRSP value-weighted index. (See Table 3 for the sample periods and the number of observations.) The predictor variables are the log dividend-price ratio ( $d - p$ ), the log earnings-price ratio ( $e - p$ ), the 3-month T-bill rate ( $r_3$ ), and the long-short yield spread ( $y - r_1$ ). The third and fourth columns report the  $t$ -statistic and the point estimate  $\hat{\beta}$  from an OLS regression of returns onto the predictor variable. The next two columns report the 90% Bonferroni confidence intervals for  $\beta$  using the  $t$ -test and  $Q$ -test, respectively. The final column reports the lower bound of the confidence interval for  $\beta$  based on the  $Q$ -test at  $\rho = 1$ .

Series	Variable	$t$ -stat	$\widehat{\beta}$	90% CI: $\beta$		Low CI $\beta$
				$t$ -test	$Q$ -test	( $\rho = 1$ )
Panel A: Full Sample						
S&P 500	$d - p$	1.977	0.093	[-0.040,0.136]	[-0.032,0.115]	-0.017
	$e - p$	2.772	0.131	[-0.002,0.190]	[0.043,0.224]	-0.023
Annual	$d - p$	2.534	0.125	[-0.007,0.178]	[0.014,0.188]	0.020
	$e - p$	2.770	0.169	[-0.009,0.240]	[0.042,0.277]	0.002
Quarterly	$d - p$	2.060	0.034	[-0.014,0.052]	[-0.009,0.044]	-0.010
	$e - p$	2.908	0.049	[-0.001,0.068]	[0.010,0.066]	0.002
Monthly	$d - p$	1.706	0.009	[-0.006,0.014]	[-0.005,0.010]	-0.005
	$e - p$	2.662	0.014	[-0.001,0.019]	[0.002,0.018]	0.001
(continued on the next page)						

Series	Variable	$t$ -stat	$\widehat{\beta}$	90% CI: $\beta$		Low CI $\beta$
				$t$ -test	$Q$ -test	( $\rho = 1$ )
Panel B: Sample through 1994						
S&P 500	$d - p$	2.243	0.141	[-0.035,0.218]	[-0.048,0.183]	-0.080
	$e - p$	3.331	0.196	[0.063,0.273]	[0.094,0.326]	-0.029
Annual	$d - p$	2.993	0.212	[0.025,0.304]	[0.056,0.332]	0.011
	$e - p$	3.409	0.279	[0.048,0.380]	[0.126,0.448]	0.012
Quarterly	$d - p$	2.304	0.053	[-0.004,0.083]	[-0.006,0.076]	-0.027
	$e - p$	3.506	0.079	[0.018,0.107]	[0.027,0.109]	0.005
Monthly	$d - p$	1.790	0.013	[-0.004,0.022]	[-0.007,0.017]	-0.013
	$e - p$	3.185	0.022	[0.002,0.030]	[0.005,0.028]	0.000
Panel C: Sample from 1952						
Annual	$d - p$	2.289	0.124	[-0.023,0.178]	[-0.007,0.183]	0.020
	$e - p$	1.733	0.114	[-0.078,0.178]	[-0.031,0.229]	-0.025
	$r_3$	-1.143	-0.095	[-0.229,0.045]	[-0.231,0.042]	—
	$y - r_1$	1.124	0.136	[-0.087,0.324]	[-0.075,0.359]	-0.156
Quarterly	$d - p$	2.236	0.036	[-0.011,0.051]	[-0.010,0.030]	0.005
	$e - p$	1.777	0.029	[-0.019,0.044]	[-0.012,0.042]	-0.003
	$r_3$	-1.766	-0.042	[-0.084,-0.004]	[-0.084,-0.004]	-0.086
	$y - r_1$	1.991	0.090	[0.009,0.162]	[0.006,0.158]	-0.002
Monthly	$d - p$	2.259	0.012	[-0.004,0.017]	[-0.004,0.010]	0.001
	$e - p$	1.754	0.009	[-0.006,0.014]	[-0.004,0.012]	-0.001
	$r_3$	-2.431	-0.017	[-0.030,-0.006]	[-0.030,-0.006]	-0.030
	$y - r_1$	2.963	0.047	[0.020,0.072]	[0.020,0.072]	0.016

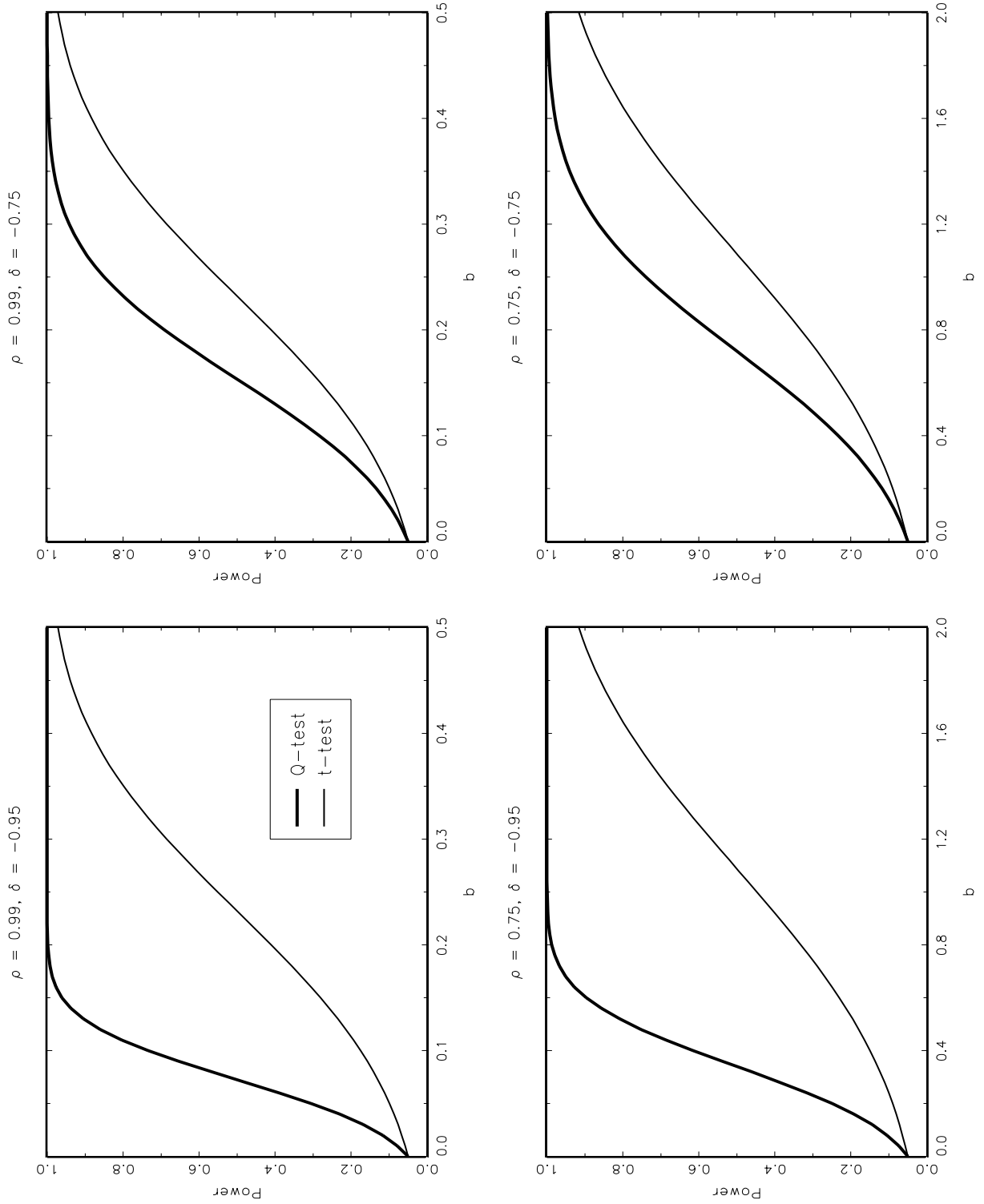


Figure 1: **Local Asymptotic Power under First-Order Asymptotics.** This figure plots the power of the  $Q$ -test and the  $t$ -test when the predictor variable is an AR(1). The null hypothesis is  $\beta = \beta_0$  against the local alternatives  $b = \sqrt{T}(\beta - \beta_0) > 0$ .  $\rho = 0.99, 0.75$  is the autoregressive root of the predictor variable, and  $\delta = -0.95, -0.75$  is the correlation between the innovations to returns and the predictor variable.

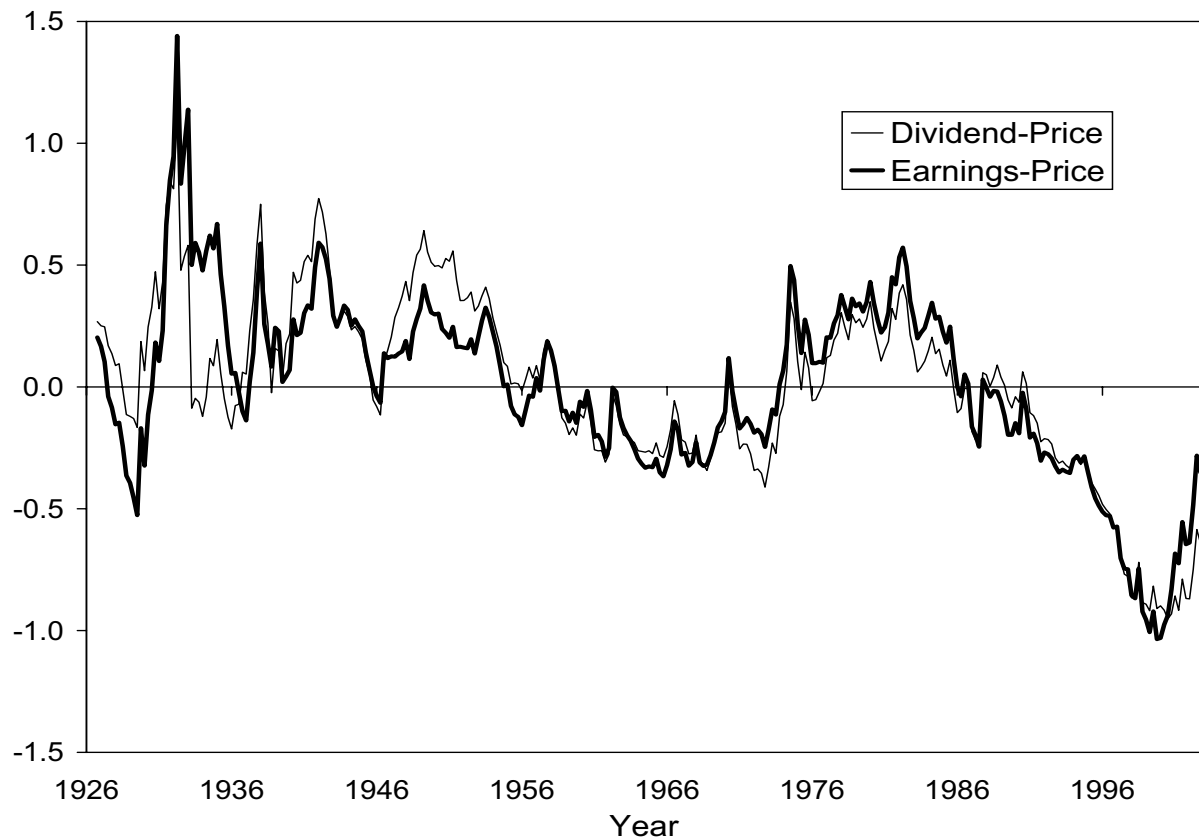


Figure 2: **Time Series Plot of the Valuation Ratios.** This figure plots the log dividend-price ratio for the CRSP value-weighted index and the log earnings-price ratio for the S&P 500. Earnings are smoothed by taking a ten-year moving average. The sample period is 1926:4–2002:4.



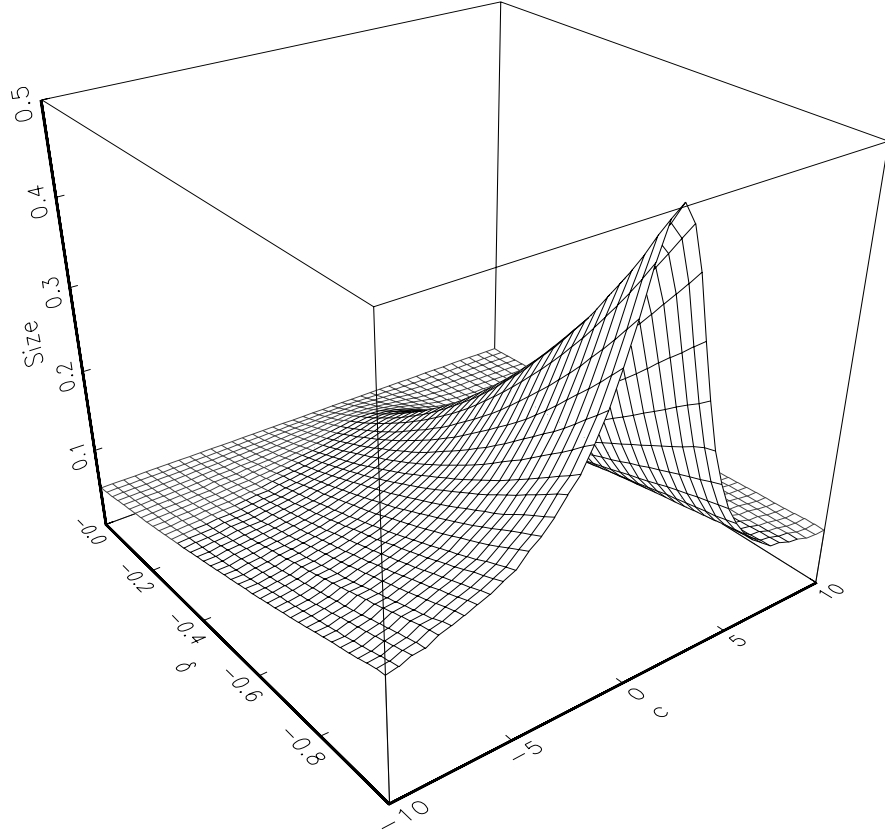


Figure 3: **Asymptotic Size of the One-Sided  $t$ -test at 5% Significance.** This figure plots the actual size of the nominal 5%  $t$ -test when the largest autoregressive root of the predictor variable is  $\rho = 1 + c/T$ . The null hypothesis is  $\beta = \beta_0$  against the one-sided alternative  $\beta > \beta_0$ .  $\delta$  is the correlation between the innovations to returns and the predictor variable. The dark shade indicates regions where the size is greater than 7.5%.

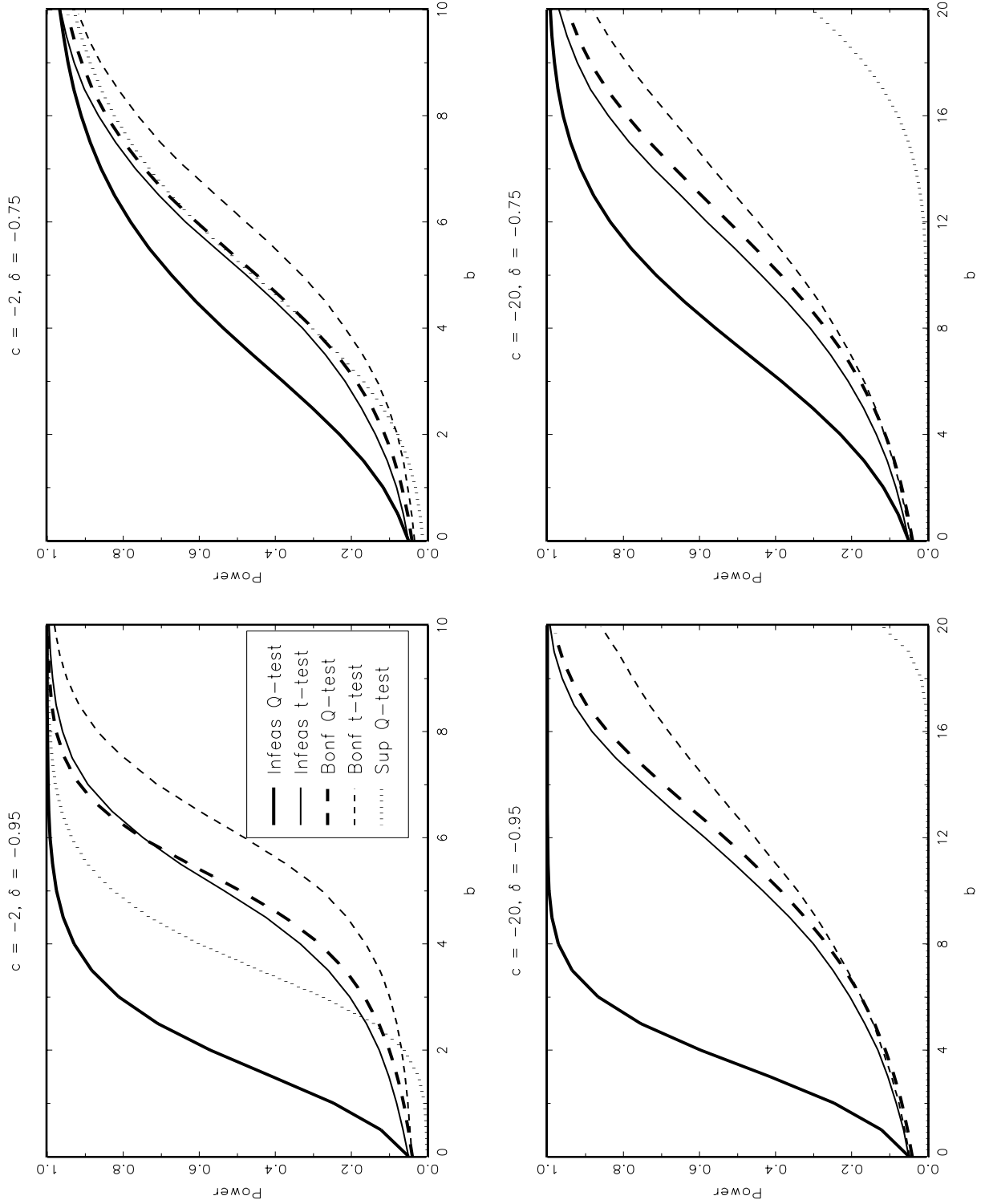


Figure 4: **Local Asymptotic Power under Local-to-Unity Asymptotics.** This figure plots the power of the infeasible  $Q$ -test and  $t$ -test that assume knowledge of the local-to-unity parameter, the Bonferroni  $Q$ -test and  $t$ -test, and the sup-bound  $Q$ -test. The null hypothesis is  $\beta = \beta_0$  against the local alternatives  $b = T(\beta - \beta_0) > 0$ .  $c = -2, -20$  is the local-to-unity parameter, and  $\delta = -0.95, -0.75$  is the correlation between the innovations to returns and the predictor variable.

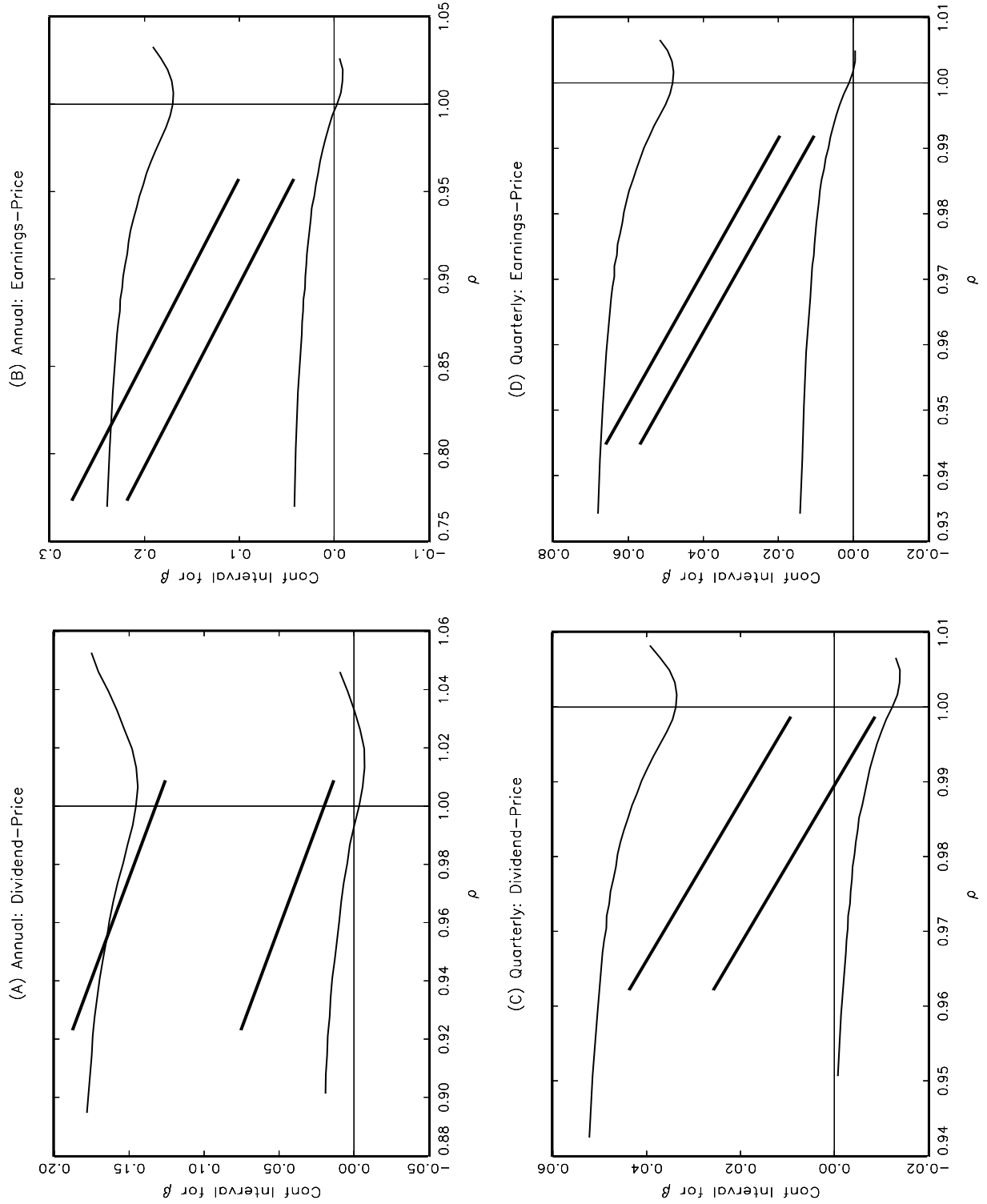


Figure 5: **Bonferroni Confidence Interval for the Valuation Ratios.** This figure plots the 90% confidence interval for  $\beta$  over the confidence interval for  $\rho$ . The significance level for  $\rho$  is chosen to result in a 90% Bonferroni confidence interval for  $\beta$ . The thick (thin) line is the confidence interval for  $\beta$  computed by inverting the  $Q$ -test ( $t$ -test). Returns are for annual and quarterly CRSP value-weighted index (1926–2002). The predictor variables are the log dividend-price ratio and the log earnings-price ratio.

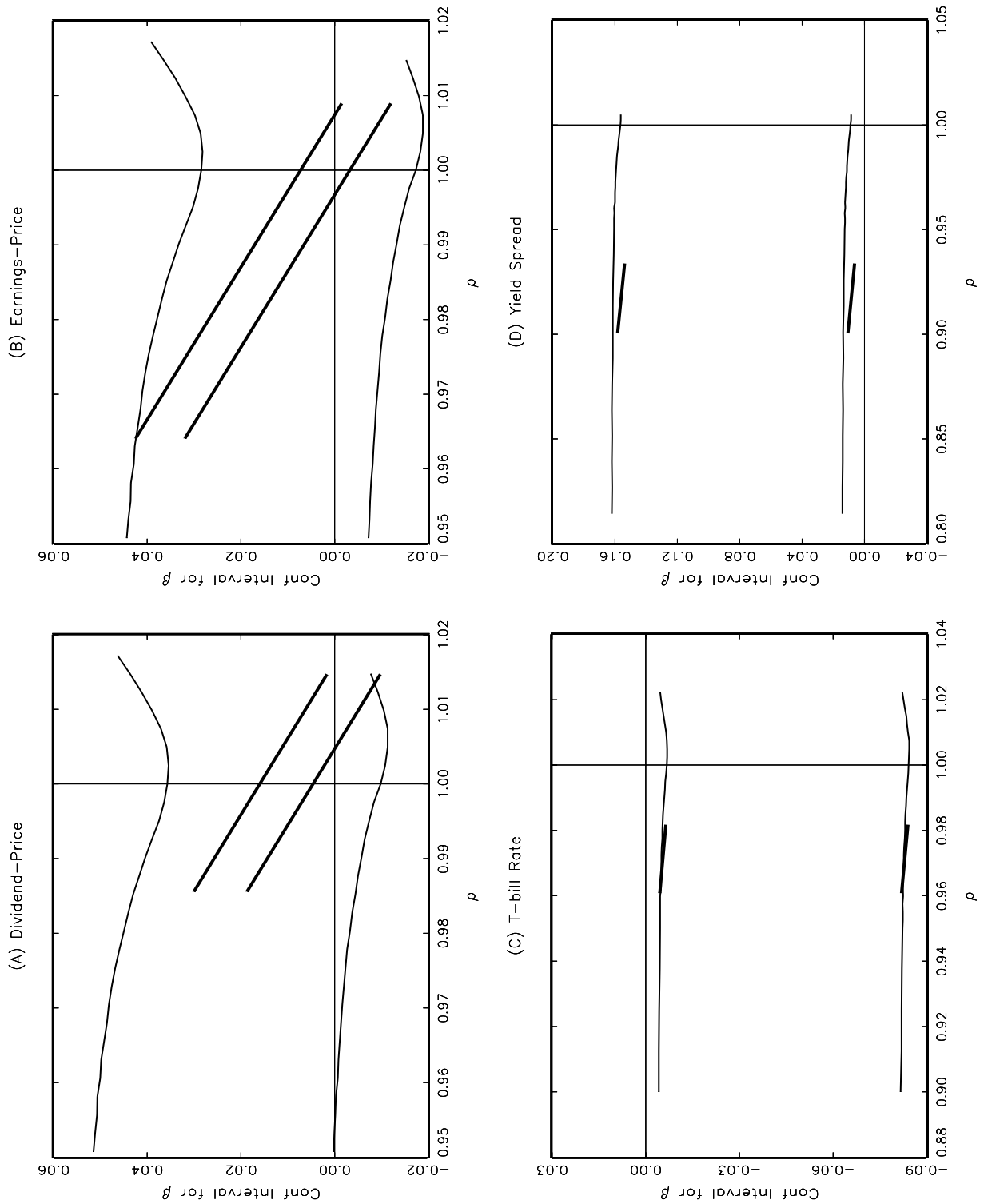


Figure 6: **Bonferroni Confidence Interval for the Sample from 1952.** This figure plots the 90% confidence interval for  $\beta$  over the confidence interval for  $\rho$ . The significance level for  $\rho$  is chosen to result in a 90% Bonferroni confidence interval for  $\beta$ . The thick (thin) line is the confidence interval for  $\beta$  computed by inverting the  $Q$ -test ( $t$ -test). Returns are for quarterly CRSP value-weighted index (1952–2002). The predictor variables are the log dividend-price ratio, the log earnings-price ratio, the 3-month T-bill rate, and the long-short yield spread.