Efficient Tests of Stock Return Predictability

John Y. Campbell and Motohiro Yogo*

October 23, 2002

Abstract

Empirical studies have suggested that stock returns can be predicted by financial variables such as the dividend-price ratio. However, these studies typically ignore the high persistence of predictor variables, which can make first-order asymptotics a poor approximation in finite samples. Using a more accurate asymptotic approximation, we propose two methods to deal with the persistence problem. First, we develop a pretest that determines when the conventional t-test for predictability is misleading. Second, we develop a new test of predictability that results in correct inference regardless of the degree of persistence and is efficient compared to existing methods. Applying our methods to US data, we find that the dividend-price ratio and the smoothed earnings-price ratio are sufficiently persistent for conventional inference to be highly misleading. However, we find some 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.

*Department of Economics, Harvard University and NBER (john_campbell@harvard.edu); and Department of Economics, Harvard University (yogo@fas.harvard.edu). First draft: June 26, 2002. For helpful comments we thank Andrew Ang, Markku Lanne, Marcelo Moreira, Robert Shiller, James Stock, Mark Watson, and participants of the 2002 Econometric Society Australasian Meeting and the Harvard Finance Workshop.

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, we 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. Hence, an important question is whether the large sample theory provides an accurate approximation to the actual finite sample distribution of the t-statistic. Unfortunately, this may not be the case since financial variables typically used as regressors tend to be highly persistent.

To be concrete, suppose we were to use the log dividend-price ratio as the regressor. Even if we were to know with certainty that the log dividend-price ratio is stationary, a time series plot (or more formally a unit root test) tells us 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% for plausible parameter values in the dividend-price ratio regression.[1] They derive an alternative asymptotic distribution theory in which the regressor is modeled as having a local-to-unit root. This theory provides a more accurate approximation to the finite sample distribution.

The issue of persistence suggests that the "significant" t-statistics in the empirical finance literature might be a consequence of size distortion rather than predictability of stock

[1] We report their result for the 10% one-sided t-test 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.

returns. Some recent papers have therefore proposed and applied test procedures that have the correct size even if the predictor variable is highly persistent or contains a unit root. For instance, 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. Using a conservative test procedure, Lanne (2002) finds no evidence that stock returns can be predicted by a highly persistent predictor variable.

A difficulty with understanding the rather large literature on predictability is the sheer variety of test procedures that have been proposed. The main contribution of this paper is to analyze tests of predictability within the unifying framework of statistical optimality theory. Using the Neyman-Pearson Lemma and local-to-unity asymptotics, we derive the Gaussian power envelope when the degree of persistence of the predictor variable is known. We show that there is no uniformly most powerful (UMP) test even asymptotically since the optimal test statistic is a weighted sum of two minimal sufficient statistics. In particular, the t-test (with appropriate critical values) fails to achieve the power envelope. However, since one of the two sufficient statistics is ancillary, there is a conditional test that is optimal and whose power function is never far below the power envelope for point optimal tests. Using the optimal conditional test, we propose a new test procedure, closely related to Lewellen (2002), that has good power and is computationally simple.

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 innovations to the predictor variable are correlated with innovations to stock returns, we can subtract a multiple of the innovation to the predictor variable from the stock return 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 has a near-unit root, it is possible to do this in a way that retains power advantages over the t-test.

Although tests derived under local-to-unity asymptotics — Cavanagh, Elliott, and Stock (1995), Lanne (2002), or the test proposed in this paper — always lead to correct inference, they are 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. In this paper, we develop a simple pretest based on the confidence interval for the largest autoregressive root. If the confidence interval indicates that the predictor variable is sufficiently stationary, one can proceed with inference based on the t-test with conventional critical values. The pretest thus provides a bridge between tests based on first-order and local-to-unity asymptotics.

We apply our methods to annual, quarterly, and monthly US data, looking first at dividend-price and 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. The dividend-price ratio also predicts returns at annual frequency, but we cannot reject the null hypothesis at quarterly and monthly frequencies.

In a 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 have prior knowledge that the largest root is non-explosive.

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, and we find strong evidence that they predict returns.

The rest of the paper is organized as follows. In Section 2, we review the theory of optimal tests of predictability under first-order asymptotics. First-order asymptotics is appropriate only when the predictor variable is sufficiently stationary. Although these results are not applicable for testing predictability with persistent financial variables, the section provides a review of the relevant statistical tools in a familiar framework. In Section 3, we derive the theory of optimal tests in the local-to-unity framework and discuss how these tests can be implemented in practice. 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.

2 Predictive Regressions

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 statistical model that we consider is

$$r_t = \beta x_{t-1} + u_t, \tag{1}$$

$$x_t = \rho x_{t-1} + v_t, \tag{2}$$

where β 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 $(u_t, v_t)'$ is independently and identically distributed (i.i.d.) with mean zero and covariance matrix $\Sigma = [(1,\delta)', (\delta, 1)']$. We further assume that the correlation δ between the innovations is known. We will later relax these assumptions to a more realistic statistical model. For now, this simple model captures the essence of the problem.

In equation (2), ρ is the unknown degree of persistence in the variable x_t. If $|\rho| < 1$ and fixed, x_t is integrated of order zero (I(0)). If $\rho = 1$, x_t is integrated of order one (I(1)). Since β and ρ are the only unknown parameters in the model, we can write down the joint log likelihood[2]

$$L(\beta, \rho) = -\sum_{t=1}^{T}[(r_t - \beta x_{t-1})^2 - 2\delta(r_t - \beta x_{t-1})(x_t - \rho x_{t-1}) + (x_t - \rho x_{t-1})^2]. \tag{3}$$

Suppose we are interested in testing the null hypothesis $\beta = \beta_0$. We consider two alternative hypotheses. The first is the simple alternative $\beta = \beta_1$, and the second is the composite alternative $\beta \neq \beta_0$. The hypothesis testing problem is complicated by the fact that the nuisance parameter ρ is unknown.

One way to test the hypothesis of interest in presence of the nuisance parameter ρ is through the likelihood ratio test (LRT). Let $\widehat{\beta}$ and $\widehat{\rho}$ denote the OLS estimators of β and ρ,

[2]To simplify notation, we ignore additive and multiplicative constants in expressions involving the likelihood throughout the paper.

respectively. Define $\widehat{\rho}(\beta) = \widehat{\rho} - \delta(\widehat{\beta} - \beta)$. Against the simple alternative, the LRT rejects the null hypothesis if

$$L(\beta_1, \widehat{\rho}(\beta_1)) - L(\beta_0, \widehat{\rho}(\beta_0)) = 2(\beta_1 - \beta_0) \sum_{t=1}^{T} x_{t-1} r_t - (\beta_1^2 - \beta_0^2) \sum_{t=1}^{T} x_{t-1}^2 > C \tag{4}$$

for some constant C. (With a slight abuse of notation, we use C to denote a generic constant throughout the paper.) Against the composite alternative, the LRT rejects the null if

$$L(\widehat{\beta}, \widehat{\rho}(\widehat{\beta})) - L(\beta_0, \widehat{\rho}(\beta_0)) = t(\beta_0)^2 > C, \tag{5}$$

where $t(\beta_0) = (\sum_{t=1}^{T} x_{t-1}^2)^{1/2}(\widehat{\beta} - \beta_0)$. In other words, the LRT against the composite alternative is based on the t-statistic.

Note that we would obtain the same tests (4) and (5) starting from the marginal likelihood $L(\beta) = -\sum_{t=1}^{T}(r_t - \beta x_{t-1})^2$. Hence, the LRT can be interpreted as a test that ignores information contained in equation (2) of the statistical model. Intuitively, this seems to be a reasonable solution to the hypothesis testing problem when ρ is unknown.

The problem with the nuisance parameter could also be resolved if ρ were known *a priori*. Since β is then the only unknown parameter in the likelihood function (3), the Neyman-Pearson Lemma implies that the most powerful test against the simple alternative rejects the null if

$$L(\beta_1, \rho) - L(\beta_0, \rho) = 2(\beta_1 - \beta_0) \sum_{t=1}^{T} x_{t-1}[r_t - \delta(x_t - \rho x_{t-1})] - (\beta_1^2 - \beta_0^2) \sum_{t=1}^{T} x_{t-1}^2 > C. \tag{6}$$

Since the optimal test statistic (6) is a weighted sum of two minimal sufficient statistics with the weights depending on the alternative β_1, there is no UMP test. However, these point optimal tests have power against the composite alternative for appropriately chosen values of β_1.

Comparing the LRT (4) with the most powerful test when ρ is known (6), they are equivalent if and only if $\delta = 0$. As noted by Torous, Valkanov, and Yan (2001) and Lewellen (2002), incorporating knowledge of ρ, if it were known, could result in large efficiency gains.

2.1 Local Asymptotic Power

Suppose x_t is I(0). Then conventional first-order asymptotics applies, and the OLS estimator $\widehat{\beta}$ 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. Formally, we consider a sequence of alternatives of the form $\beta_1 = \beta_0 + \bar{b}/\sqrt{T}$ for some fixed constant $\bar{b}$. (See Lehmann (1999, Chapter 3) for a textbook treatment of local alternatives and relative efficiency.)

The LRT against the simple alternative (4) rejects the null if

$$L(\beta_1, \widehat{\rho}(\beta_1)) - L(\beta_0, \widehat{\rho}(\beta_0)) = 2\bar{b}T^{-1/2}\sum_{t=1}^{T} x_{t-1}(r_t - \beta_0 x_{t-1}) - \bar{b}^2 T^{-1}\sum_{t=1}^{T} x_{t-1}^2 > C. \tag{7}$$

The test statistic is a weighted sum of two statistics, $T^{-1/2}\sum_{t=1}^{T} x_{t-1}(r_t - \beta_0 x_{t-1})$ and $T^{-1}\sum_{t=1}^{T} x_{t-1}^2$, with the weights depending on $\bar{b}$. However, the second statistic has a degenerate asymptotic distribution (i.e. it converges in probability to $\mathbf{E}[x_{t-1}^2] = \sigma_x^2 = (1-\rho^2)^{-1}$). Hence, the LRT is asymptotically equivalent to the test that rejects if

$$\frac{\bar{b}\sum_{t=1}^{T} x_{t-1}(r_t - \beta_0 x_{t-1})}{(\sum_{t=1}^{T} x_{t-1}^2)^{1/2}} = \bar{b}t(\beta_0) > C. \tag{8}$$

Since the test takes the same form for each alternative $\bar{b}$, the t-test is asymptotically the LRT against the simple alternative.

Now consider the case that ρ is known. Then the most powerful test against the simple alternative (6) rejects if

$$L(\beta_1, \rho) - L(\beta_0, \rho) = 2\bar{b}T^{-1/2}\sum_{t=1}^{T} x_{t-1}[r_t - \beta_0 x_{t-1} - \delta(x_t - \rho x_{t-1})] - \bar{b}^2 T^{-1}\sum_{t=1}^{T} x_{t-1}^2 > C. \tag{9}$$

The optimal test statistic is a weighted sum to two minimal sufficient statistics with the weights depending on $\bar{b}$. However, the second statistic $T^{-1}\sum_{t=1}^{T} x_{t-1}^2$ has a degenerate asymptotic distribution, so the most powerful test is asymptotically equivalent to the test that rejects if

$$\frac{\bar{b}\sum_{t=1}^{T} x_{t-1}[r_t - \beta_0 x_{t-1} - \delta(x_t - \rho x_{t-1})]}{(1-\delta^2)^{1/2}(\sum_{t=1}^{T} x_{t-1}^2)^{1/2}} = \bar{b}Q(\beta_0, \rho) > C, \tag{10}$$

where

$$Q(\beta_0, \rho) = \frac{(\sum_{t=1}^{T} x_{t-1}^2)^{1/2}[(\widehat{\beta} - \beta_0) - \delta(\widehat{\rho} - \rho)]}{(1-\delta^2)^{1/2}}. \tag{11}$$

Since the test based on $Q(\beta_0, \rho)$ takes the same form for each alternative $\bar{b}$, it is UMP against one-sided alternatives when ρ is known.

When $\beta_0 = 0$, the statistic $Q(\beta_0, \rho)$ is the t-statistic that results from regressing $r_t - \delta(x_t - \rho x_{t-1})$ onto x_{t-1}. Since $v_t = x_t - \rho x_{t-1}$, knowledge of ρ 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. The statistic $Q(\beta_0, \rho)$ has also appeared in Lewellen (2002), where he motivates it by interpreting $\delta(\widehat{\rho} - \rho)$ in (11) as the "finite sample bias" of the OLS estimator (cf. Stambaugh (1999)). Using the fact that $Q(\beta_0, \rho)$ is asymptotically standard normal under the null and assuming that $\rho = 1$, Lewellen (2002) tests the predictability of returns using the statistic $Q(\beta_0, 1)$. We have shown here that the UMP one-sided test when ρ is known is based on the statistic (11) rather than the conventional t-statistic. For simplicity, we will refer to this (infeasible) test as the Q-test.

2.2 Power under First-Order Asymptotics

We now derive the power functions of the t-test and the Q-test to illustrate the power gains that would result from incorporating knowledge of the persistence parameter ρ. Let $\Phi(z)$ denote one minus the cumulative distribution function of the standard normal, and let z_α denote the upper α-quantile of that distribution. Under first-order asymptotics, the probability of rejecting a local alternative $b = \bar{b}$ is

$$\pi_t(\bar{b}) = \Phi(z_\alpha - \sigma_x|\bar{b}|), \tag{12}$$

$$\pi_Q(\bar{b}) = \Phi\left(z_\alpha - \frac{\sigma_x|\bar{b}|}{(1-\delta^2)^{1/2}}\right), \tag{13}$$

for the t-test and the Q-test, respectively. Since the Q-test is asymptotically UMP against one-sided alternatives, $\pi_Q(b)$ is also the power envelope when ρ is known.

In Figure 1, we plot the power functions for various combinations of ρ (0.99 and 0.75) and δ (-0.95 and -0.75). These values are chosen to correspond to the relevant region of the parameter space when the predictor variable is the log dividend-price ratio or the log earnings-price ratio. Note that, as expected, the power function for the Q-test dominates that of the t-test. A comparison of (12) and (13) shows that the power gain arises from $\delta^2 \neq 0$ and is increasing in the degree of correlation. Intuitively, when ρ is known, the innovation $v_t = x_t - \rho x_{t-1}$ is known as well. Then by subtracting off the portion of the innovation to r_t that is correlated with v_t (i.e. δv_t), the Q-test is able to gain efficiency from the reduction in noise. When the predictor variable is a valuation ratio (e.g. dividend-price ratio or earnings-price ratio), the efficiency 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 ρ is unknown, so unfortunately, the large efficiency gains over the t-test cannot be realized.

2.3 Relaxing the Assumptions

The statistical model (1)–(2) and the distributional assumptions that we have used to derive the results in the last section are quite restrictive. In this section, we show that all the key insights are retained under a more general model. Consider the statistical model

$$r_t = \gamma_r + \beta x_{t-1} + u_t, \tag{14}$$

$$x_t = \gamma_x + \rho x_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i} + v_t. \tag{15}$$

The predictor variable x_t is now an AR(p), which we have written in the augmented Dickey-Fuller form. x_t is I(0) if $\rho < 1$ and fixed and is I(1) if $\rho = 1$. We assume that all the other roots ψ_i $(i = 1, \ldots, p-1)$ are fixed and less than one in absolute value.

Following Elliott and Stock (1994), we make the following fairly weak distributional assumptions:

Assumption 1 *Let* $w_t = (u_t, v_t)'$ *and* $\mathcal{F}_t = \{w_s | s \leq t\}$ *be the filtration generated by the process* w_t*. Then*

1. $\mathbf{E}[w_t | \mathcal{F}_{t-1}] = 0$,
2. $\mathbf{E}[w_t w_t' | \mathcal{F}_{t-1}] = \Sigma = [(\sigma_u^2, \sigma_{uv})', (\sigma_{uv}, \sigma_v^2)']$,
3. $\sup_t \mathbf{E}[u_t^4] < \infty$ *and* $\sup_t \mathbf{E}[v_t^4] < \infty$.

In other words, w_t is a homoskedastic martingale difference sequence with finite fourth moments. Under this assumption, the asymptotics for the t-statistic and the Q-statistic continue to hold through the law of large numbers and the central limit theorem.

For instance, consider the Q-statistic

$$Q(\beta_0, \rho) = \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[r_t - \beta_0 x_{t-1} - (\sigma_{uv}/\sigma_v^2)(x_t - \rho x_{t-1} - \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i})]}{\sigma_u (1-\delta^2)^{1/2} (\sum_{t=1}^{T} x_{t-1}^{\mu 2})^{1/2}}, \tag{16}$$

where $x_{t-1}^{\mu} = x_{t-1} - T^{-1} \sum_{t=1}^{T} x_{t-1}$. Definition (16) is a generalization of definition (11) to the model (14)–(15). Suppose ρ is known. The Q-statistic (16) is still infeasible because it requires knowledge of nuisance parameters Σ and ψ_i. However, a feasible version of the Q-statistic that replaces these nuisance parameters with consistent estimators has the same asymptotic distribution as the infeasible Q-statistic. Hence, the power function for the Q-statistic in Figure 1 still applies as long as the largest autoregressive root ρ is known. The asymptotic results that we have derived for the simplified statistical model (1)–(2) therefore carry over to the more general model (14)–(15).

Under the further assumption that w_t is bivariate normal, we recover the result that the power function for the Q-test coincides with the power envelope for point optimal tests. When w_t is non-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. This illustrates the fact that the Gaussian likelihood function and the Neyman-Pearson Lemma can be useful tools for deriving efficient tests even if the error distribution is unknown.

3 Inference with a Persistent Regressor

In Figure 2, we plot the log dividend-price ratio for the CRSP NYSE/AMEX portfolio and the log smoothed earnings-price ratio for the S&P 500 portfolio 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 ρ are $[0.964, 1.010]$ and $[0.949, 1.005]$ for the dividend-price ratio and the earnings-price ratio, respectively. Hence, we cannot reject the null hypothesis that these valuation ratios contain a unit root.

The persistence of financial variables typically used to predict returns has important implications for inference about predictability. Even if x_t is I(0), first-order asymptotics is a poor approximation in finite samples as long as ρ is close to one because of the discontinuity in the asymptotic distribution at $\rho = 1$ (cf. Elliott and Stock (1994)). Local-to-unity asymptotics is an alternative asymptotic framework that circumvents this problem by modeling $\rho = 1 + c/T$, where c is 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. $c < 0$). Local-to-unity asymptotics 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.) The local-to-unity framework has been applied to the present context of predictive regressions by several authors. Elliott and Stock (1994) derived the asymptotic null distribution of the t-statistic. This has been extended to long-horizon t-tests by Torous, Valkanov, and Yan (2001).

An important feature of the nearly integrated case is that the mean of the process x_t is not well defined. Hence, the process x_t and the demeaned process $x_t^\mu = x_t - T^{-1}\sum_{t=1}^T x_t$ have different asymptotic distributions. Similarly, second moments do not exist for nearly integrated series. However, when appropriately scaled, these objects converge to functionals of a diffusion process. Let $J_c(s)$ be the diffusion process defined by the stochastic differential equation $dJ_c(s) = cJ_c(s)ds + dW(s)$ with initial condition $J_c(0) = 0$, where $W(s)$ is a Wiener process. 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 on the space $D[0, 1]$ of cadlag functions (cf. Billingsley (1999, Chapter 3)). By a straightforward extension of Phillips (1987, Lemma 1),

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

where $\omega = (1 - \sum_{i=1}^{p-1}\psi_i)^{-1}\sigma_v$.

In empirical application, the series x_t needs to be demeaned. 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. Hence, it is natural to have intercept terms as in the model (14)–(15), which were assumed away in the simplified model (1)–(2). Throughout the rest of the paper, we assume that (14) and (15) are the true processes for excess returns and the predictor variable, respectively, where $c = T(\rho - 1)$

is fixed. We restrict ourselves to tests that are invariant to translations in the unknown intercept terms γ_r and γ_x (cf. Lehmann (1986, Chapter 6)).

3.1 Point Optimal Tests

In this section, we derive optimal test procedures in the local-to-unity framework, mirroring our derivations for conventional first-order asymptotics in Section 2.1. In order to do so, we strengthen Assumption 1 and assume the following

Assumption 2

1. *w_t is independently distributed* $\mathbf{N}(0, \Sigma)$.

2. *The nuisance parameters Σ and ψ_i $(i = 1, \ldots, p-1)$ are known.*

We will later relax these assumptions and show that the asymptotic results hold more generally.

Since the derivations of the LRT (4) and the point optimal test (6) against the simple alternative did not rely on assumptions about the nature of ρ, they are still applicable here. However, to derive expressions analogous to (7) and (9) for tests against a local alternative, we must consider alternatives that are in a T^{-1}-neighborhood of β_0. This is because when the regressor x_t contains a local-to-unit root, OLS estimators $\widehat{\beta}$ and $\widehat{\rho}$ are consistent at the rate T, rather than $\sqrt{T}$. Formally, we consider a sequence of alternatives of the form $\beta_1 = \beta_0 + \overline{b}/T$ for some fixed constant $\overline{b}$.

Define the test statistic

$$N(\beta_0, \overline{b}) = 2\overline{b}T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}(r_t - \beta_0 x_{t-1}) - \overline{b}^2 T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}. \tag{17}$$

The LRT against the simple alternative rejects the null if $N(\beta_0, \overline{b}) > C$. Let $\beta = \beta_0 + b/T$ denote the true value of the unknown parameter. We show in the Appendix that

$$N(\beta_0, \overline{b}) \Rightarrow (2\overline{b}b - \overline{b}^2)\omega^2\kappa_c^2 + 2\overline{b}\sigma_u\omega[\delta\tau_c + (1-\delta^2)^{1/2}\kappa_c Z], \tag{18}$$

where $\kappa_c = (\int J_c^{\mu}(s)^2 ds)^{1/2}$, $\tau_c = \int J_c^{\mu}(s)dW(s)$, and Z is a standard normal random variable independent of $(W(s), J_c(s))$. Suppose $\overline{b} > 0$ so that for an α-level test, we reject the null if $N(\beta_0, \bar{b}) > C_\alpha$, where C_α is the upper α-quantile of $N(\beta_0, \bar{b})$ under the null (i.e. $b = 0$). Although the statistic $N(\beta_0, \bar{b})$ is the LRT against the alternative $b = \bar{b}$, expression (18) shows that it has power against all alternatives $b > 0$.

Recall that the LRT against the composite alternative is based on the t-statistic. As shown by Elliott and Stock (1994), the t-statistic has the asymptotic distribution

$$t(\beta_0) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u} + \delta\frac{\tau_c}{\kappa_c} + (1-\delta^2)^{1/2}Z. \tag{19}$$

Because $T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}$ has a non-degenerate asymptotic distribution under local-to-unity asymptotics, the LRT against the simple alternative is not asymptotically equivalent to the LRT against the composite alternative.

Now consider the case that ρ is known, or equivalently c is known. Define the test statistic

$$\begin{aligned} P(\beta_0, \bar{b}, \rho) \quad = \quad & 2\bar{b}T^{-1}\sum_{t=1}^{T} x_{t-1}^{\mu}\left[r_t - \beta_0 x_{t-1} - \frac{\sigma_{uv}}{\sigma_v^2}\left(x_t - \rho x_{t-1} - \sum_{i=1}^{p-1}\psi_i \Delta x_{t-i}\right)\right] \\ & -\bar{b}^2 T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}. \end{aligned} \tag{20}$$

The most powerful test against a local alternative rejects if $P(\beta_0, \bar{b}, \rho) > C$. There is a pair of minimal sufficient statistics for this decision problem. Unlike the case for first-order asymptotics, the second statistic $T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}$ has a non-degenerate asymptotic distribution, so there are two minimal sufficient statistics even asymptotically. Since the optimal test statistic is a weighted sum of these two statistics, where the weights depend on the alternative $\bar{b}$, we do not have a UMP test against a one-sided alternative. Instead, we have an infinite family of asymptotically admissible tests indexed by $\bar{b}$ that are optimal against a point alternative.

We show in the Appendix that

$$P(\beta_0, \bar{b}, \rho) \Rightarrow (2\bar{b}b - \bar{b}^2)\omega^2\kappa_c^2 + 2\bar{b}\sigma_u\omega(1-\delta^2)^{1/2}\kappa_c Z. \tag{21}$$

As we have argued for the LRT, the statistic $P(\beta_0, \bar{b}, \rho)$ has power against all alternatives with the same sign as $\bar{b}$. A comparison of expressions (18) and (21) shows that the cost of not knowing ρ is an extra term $2\bar{b}\sigma_u\omega\delta\tau_c$. When δ is large, the additional noise in the test statistic should translate to a decrease in power.

Although there are two minimal sufficient statistics for the point optimal test (20), the second statistic $T^{-2}\sum_{t=1}^{T} x_{t-1}^{\mu 2}$ is ancillary. That is, its distribution does not depend on β. It is thus reasonable to consider tests that condition on the ancillary statistic. The conditional test is based on the Q-statistic (16), which has the asymptotic distribution

$$Q(\beta_0, \rho) \Rightarrow \frac{b\omega\kappa_c}{\sigma_u(1-\delta^2)^{1/2}} + Z \tag{22}$$

as shown in the Appendix. Hence, $Q(\beta_0, \rho)$ is distributed standard normal under the null (i.e. $b = 0$). Lewellen (2002) derived this result under first-order asymptotics; we generalize it to the case of local-to-unity asymptotics. Note that this statistic is pivotal in the sense that its distribution under the null does not depend on the nuisance parameter c. In contrast, the asymptotic null distributions of $N(\beta_0, \bar{b})$, $P(\beta_0, \bar{b}, \rho)$, and $t(\beta_0)$ depend on the random variables κ_c and τ_c, which have nonstandard distributions that depend on c. Of course, the distribution of these random variables can be simulated by Monte Carlo, but the test based on $Q(\beta_0, \rho)$ is much more convenient to implement computationally.

Against a local alternative $\bar{b} > 0$, we reject the null if $Q(\beta_0, \rho) > z_\alpha$. The power function of the Q-test is thus given by

$$\pi_Q(b) = \mathbf{E}\left[\Phi\left(z_\alpha - \frac{\omega\kappa_c|b|}{\sigma_u(1-\delta^2)^{1/2}}\right)\right], \tag{23}$$

where expectation is taken over the distribution of κ_c. The power of the Q-test, of course, will be dominated by the power envelope since there is no UMP test. Moreover, the Q-test is not a member of the family of point optimal tests. However, since it is the most powerful test conditional on the ancillary statistic, it should have good power properties. We will examine this in the next section.

3.2 Power under Local-to-Unity Asymptotics

Under first-order asymptotics, the t-statistic is asymptotically pivotal. That is, its asymptotic null distribution does not depend of the nuisance parameter ρ. Under local-to-unity asymptotics, however, the asymptotic null distribution of the statistic $N(\beta_0, \bar{b})$ or the t-statistic depends on the unknown nuisance parameter c, making the LRT infeasible. In this section, we examine the power properties of various test procedures under the assumption that c is known. We return to the problem of feasible tests in Section 3.4.

In Figure 3, we plot the power envelope for the point optimal tests using the local-to-unity asymptotic distribution (21). We also plot the power functions for the Q-test and the t-test (using the appropriate critical value that depends on c). Under local-to-unity asymptotics, power functions are not symmetric in b. We only report results for right-tailed tests (i.e. $b > 0$) since the results are similar for left-tailed tests. We consider various combinations of c (-2 and -20) and δ (-0.95 and -0.75), which are in the relevant region of the parameter space for the log dividend-price ratio or the log earnings-price ratio. The nuisance parameters are normalized as $\sigma_u = \omega = 1$.

Although not reported in the figure, point optimal tests that are optimal for a fixed alternative $b = \bar{b}$ have good power against all alternatives b that are of the same sign as $\bar{b}$. Hence, although we do not have a UMP test, we have an infinite family of point optimal tests that are effectively UMP. This is similar to a remarkable result by Elliott, Rothenberg, and Stock (1996) that although there is no UMP test for an autoregressive unit root, there is a family of point optimal tests that in practice achieve the power envelope.

The Q-test is quite powerful, effectively achieving the power envelope, especially for alternatives that are close to the null. Although a member of the family of point optimal tests is more powerful than the Q-test in principle, the latter has some important computational advantages. The critical value of the Q-test just depends on the quantiles of the standard normal. On the other hand, one has to run Monte Carlo simulations to calculate the critical values for the point optimal tests. Also, to assure that the point optimal test achieves good power, one must compute a reasonable value for $\bar{b}$ given c and δ. Perhaps the best value is the $\bar{b}$ such that the power function is tangent to the power envelope at 50% power.[3] This requires additional expensive Monte Carlo simulations to pick an appropriate $\bar{b}$. For these reasons, the Q-test seems more practical, especially since it does a fairly good job of approximating the power envelope.

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

[3] Stock (1994) makes similar recommendations for the point optimal unit root tests.

is large, there are large efficiency gains from subtracting off the part of the innovation to returns that is correlated with the innovation to the predictor variable.

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 $\delta = -0.95$ in the upper left panel. 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 power envelope achieves 50% power when $b = 1.7$. Following the discussion in Stock (1994, p. 2775), the Pitman efficiency of the t-test relative to the Q-test is $4.8/1.7 \approx 2.8$. This means that to achieve 50% power, the t-test asymptotically requires 180% more observations than the Q-test.

As was the case for first-order asymptotics in Section 2.3, the asymptotic power functions computed in this section are valid under the more general Assumption 1. For instance, the nuisance parameters Σ and ψ can be substituted by consistent estimators without consequence to the asymptotic theory. The only fact that we lose by dropping Assumption 2 is the point optimality of tests based on the statistic $P(\beta_0, \bar{b}, \rho)$. When w_t is non-normal, one can in principle construct a more powerful test using the relevant likelihood function if the error distribution were known. In practice, the true distribution is unknown, so the quasi-likelihood approach that we have taken here is a reasonable solution.

3.3 Relation to First-Order Asymptotics and a Simple Pretest

In this section, we discuss the relation between first-order and local-to-unity asymptotics and use it to develop a simple pretest that can be used to determine when inference based on first-order asymptotics is reliable.

Recall the asymptotic distribution of the t-statistic under local-to-unity asymptotics (19). In general, the distribution under the null is nonstandard because of its dependence on τ_c and κ_c. However, the t-statistic is standard normal in the special case $\delta = 0$. Hence, the t-statistic should be approximately standard normal when $\delta \approx 0$. Likewise, the t-statistic should be approximately standard normal when $c \ll 0$ because conventional first-order asymptotics should be a good approximation when the predictor variable is stationary. This follows formally from Phillips (1987, Theorem 2) who shows that $\tau_c/\kappa_c \Rightarrow \widetilde{Z}$ as $c \to -\infty$, where $\widetilde{Z}$ is a standard normal random variable independent of Z.

In Figure 4, we plot the actual size of the nominal 5% one-sided t-test as a function of c and δ. In other words, we plot

$$p(c, \delta; \alpha) = \Pr\left(\delta \frac{\tau_c}{\kappa_c} + (1-\delta^2)^{1/2} Z > z_\alpha\right), \tag{24}$$

where $\alpha = 0.05$. The t-test that uses conventional critical values has approximately the correct size when δ is small in absolute value or c is large in absolute value.[4] 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 τ_c/κ_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. 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 use Figure 4 to tabulate the values of $c \in (\underline{c}, \overline{c})$ that cause the size of the right-tailed t-test to exceed 7.5% for selected values of δ. 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 sufficiently reliable. Suppose a researcher is willing to tolerate an actual size of up to $\widetilde{\alpha}$ (e.g. 7.5%) for a nominal α-level (e.g. 5%) test. Let $\Theta = \{c, \delta | p(c, \delta; \alpha) > \widetilde{\alpha}\}$. Then the goal is to test

$$\begin{aligned} H_0 &: \quad \{c, \delta\} \in \Theta \\ H_1 &: \quad \{c, \delta\} \notin \Theta. \end{aligned}$$

To test this hypothesis, we first construct a $100(1-\alpha_1)\%$ confidence interval for c, which we denote as $C_c(\alpha_1)$. (For instance, the confidence interval can be computed by inverting the Dickey-Fuller test as in Stock (1991).) We then estimate δ using the OLS residuals from

[4] 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.

(14) and (15). We reject the null hypothesis if $C_c(\alpha_1) \bigcap (\underline{c}, \overline{c}) = \emptyset$, where $(\underline{c}, \overline{c})$ is taken from Table 1 using the estimated correlation $\widehat{\delta}$. That is, we reject the null if the confidence interval for c indicates that the predictor variable is sufficiently away from unit root for the t-test to be reliable. Asymptotically, this pretest has size α_1.

3.4 Feasible Tests of Predictability

In Sections 3.1–3.2, we proceeded under the assumption that c is known to develop efficient tests in that context. In practice, however, c is an unknown nuisance parameter that cannot be estimated consistently. Consequently, tests based on $N(\beta_0, \overline{b})$ and $P(\beta_0, \overline{b}, \rho)$ are infeasible since their asymptotic null distributions depend on c through the random variables κ_c and τ_c. In other words, the statistics $N(\beta_0, \overline{b})$ and $P(\beta_0, \overline{b}, \rho)$ are not asymptotically pivotal. Although the statistic $Q(\beta_0, \rho)$ is asymptotically pivotal, we still require the true value of ρ (or equivalently c) to compute the test statistic.

The problem that the tests cannot be implemented without knowledge of c is not unique to these efficient tests, but rather plagues even the conventional t-test as expression (19) reveals. Intuitively, the degree of persistence, controlled by the parameter c, influences the finite sample 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 and N-test), the value of the test statistic itself (e.g. Q-test), or both (e.g. point optimal test). Cavanagh, Elliott, and Stock (1995) discuss several methods of approaching this problem including sup-bound, Bonferroni, and Scheffe-type confidence intervals that have the correct coverage. Here, we will discuss the Bonferroni confidence interval.

To construct a Bonferroni confidence interval, we first construct a $100(1-\alpha_1)\%$ confidence interval $C_\rho(\alpha_1)$ for ρ. (Note that we parameterize the degree of persistence by ρ rather than c since this is the more natural choice in the following.) Then for each value of ρ in the confidence interval, we construct a $100(1-\alpha_2)\%$ confidence interval $C_{\beta|\rho}(\alpha_2)$ for β given ρ. A confidence interval that does not depend on ρ 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 $C_\beta(\alpha)$ is likely to be greater than $100(1-\alpha)\%$. To see this, we use 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 not known, the Bonferroni confidence interval 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 ρ. 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 Dickey-Fuller t-statistic to first construct $C_\rho(\alpha_1)$. They then suggest inverting the conventional t-statistic for testing β, using appropriate critical values computed by its asymptotic distribution (19). The two t-statistics are correlated, which tends to increase the coverage rate of the confidence interval. Cavanagh, Elliott, and Stock suggest adjusting α_1 and α_2 to achieve an exact test of the desired significance level. The method that we have outlined here has been applied to US data by Torous, Valkanov, and Yan (2001).

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 (cf. Elliott, Rothenberg, and Stock (1996)), there is no uniformly most accurate confidence interval for ρ. However, as discussed in Elliott and Stock (2001), inverting a relatively efficient unit root test translates to a relatively tight confidence interval. Hence, inverting the DF-GLS test of Elliott, Rothenberg, and Stock (1996) should result in a tighter confidence interval for ρ than inverting the Dickey-Fuller t-test. Hence, we will construct the confidence interval for ρ by applying Stock's (1991) method of confidence belts to the DF-GLS test.

In addition, the power calculations in Section 3.2 suggest that there are tests of β given ρ that are more powerful than the t-test. In particular, we can obtain a more accurate confidence interval $C_{\beta|\rho}(\alpha_2)$ by inverting the Q-test. Because the statistic $Q(\beta_0, \rho)$ is standard normal under the null, an equal-tailed α_2-level confidence interval is simply $C_{\beta|\rho}(\alpha_2) = [\underline{\beta}(\rho, \alpha_2), \overline{\beta}(\rho, \alpha_2)]$ where

$$\begin{aligned}
\widehat{\beta}(\rho) &= \frac{\sum_{t=1}^{T} x_{t-1}^{\mu}[r_t - (\sigma_{uv}/\sigma_v^2)(x_t - \rho x_{t-1} - \sum_{i=1}^{p-1} \psi_i \Delta x_{t-i})]}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}}, && (25)\\
\underline{\beta}(\rho, \alpha_2) &= \widehat{\beta}(\rho) - z_{\alpha_2/2}\sigma_u \left(\frac{1-\delta^2}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}}\right)^{1/2}, && (26)\\
\overline{\beta}(\rho, \alpha_2) &= \widehat{\beta}(\rho) + z_{\alpha_2/2}\sigma_u \left(\frac{1-\delta^2}{\sum_{t=1}^{T} x_{t-1}^{\mu 2}}\right)^{1/2}. && (27)
\end{aligned}$$

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

$$C_\beta(\alpha) = \begin{cases} \left[\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2), \overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2)\right] & \text{if } \delta < 0 \\ \left[\underline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2), \overline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2)\right] & \text{otherwise.} \end{cases} \tag{28}$$

Hence, we have a closed form expression for the confidence interval of β 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 α_1 and α_2 can be adjusted to achieve a test of desired significance level $\widetilde{\alpha} \leq \alpha$. To do so, we first fix α_2. Then for each $\delta < 0$, we numerically search over a grid for c to find the $\overline{\alpha}_1$ such that

$$\Pr(\underline{\beta}(\overline{\rho}(\overline{\alpha}_1), \alpha_2) > \beta) \leq \widetilde{\alpha}/2, \tag{29}$$

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

$$\Pr(\overline{\beta}(\underline{\rho}(\underline{\alpha}_1), \alpha_2) < \beta) \leq \widetilde{\alpha}/2. \tag{30}$$

In Table 2, we report the values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ for selected values of δ when $\widetilde{\alpha} = \alpha_2 = 0.10$, using the grid $c \in [-50, 5]$. The table can be used to construct a 5% one-sided Q-test for predictability. Note that $\underline{\alpha}_1$ and $\overline{\alpha}_1$ are increasing in δ, so the Bonferroni inequality has more slack and the unadjusted Bonferroni test is more conservative the smaller is δ in absolute value.

Our computational results indicate that in general the inequalities (29) and (30) are close to equalities when c is large and are slack when c is small. For right-tailed tests, the probability (29) can be as small as 0.04 for some values of c and δ. For left-tailed tests, the probability (30) can be as small as 0.012. This means that even the adjusted Bonferroni Q-test is still conservative (i.e. undersized) when $c < 5$. In principle, one can obtain a tighter Bonferroni confidence interval for β by using confidence belts that are narrower than the DF-GLS confidence belt for stationary autoregressive roots and wider for explosive roots. The approach that we have taken here is somewhat conservative but tractable. Similar tests that have size exactly $\widetilde{\alpha}$ uniformly in c are elusive and are left to future research.

3.5 Power of Feasible Tests

In this section, we analyze the power properties of various feasible tests that have been proposed in the literature.

In addition to the Bonferroni Q-test described in the last section, we analyze the Bonferroni t-test. Our Bonferroni t-test is a slight modification of the one originally proposed by Cavanagh, Elliott, and Stock (1995); instead of constructing the confidence interval for c using the Dickey-Fuller t-test, we use the DF-GLS test of Elliott, Rothenberg, and Stock (1996). We use the numerical procedure described in the last section to set the size of the Bonferroni t-test test to 5% uniformly in $c \in [-50, 5]$.

In Figure 5, we plot the power of the two Bonferroni tests against right-sided local alternatives (i.e. $b > 0$). As a benchmark, we also plot the power function of the infeasible t-test that assumes knowledge of c. The values of c (-2 and -20) and δ (-0.95 and -0.75) are the same as those in Figure 3.

When $c = -2$, the Bonferroni Q-test dominates the Bonferroni t-test. The Bonferroni Q-test comes very close to the power function of the infeasible t-test. When $c = -2$ and $\delta = -0.95$, the Pitman efficiency of the Bonferroni Q-test over the Bonferroni t-test is 1.2, which means that the t-test requires 20% more observations to achieve 50% power. When $c = -20$, both tests have similar power with the power functions lying slightly below that of the infeasible t-test. This is not surprising since the t-statistic is approximately pivotal when $c \ll 0$, so the power loss from not knowing c is relatively small.

In addition to the the Bonferroni tests, we also consider the power of Lewellen's (2002) test which is the Q-test that assumes $\rho = 1$. In our notation (28), 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 ρ 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 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 3.3 that the left-tailed t-test is undersized when $\delta < 0$.) 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 predictor variable is stationary (i.e. $c = -20$).

We conclude that the Bonferroni Q-test has important power advantages over the other feasible tests. Against right-sided alternatives, it has greater power than the Bonferroni t-test when the predictor variable is highly persistent, and it has much greater 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 excess stock returns, dividend-price ratio, and earnings-price ratio. The first is annual S&P 500 index data (1871–2001) from DRI-WEFA Webstract since 1926 and Shiller (2000) before then.[5] The last three are annual, quarterly, and monthly NYSE/AMEX value-weighted index data (1926–2001) 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 monthly frequency only since 1935, so we use Shiller's annual earnings before then. Instead of using linear extrapolation of annual earnings as in Shiller (2000), we assign annual earnings to each month of the year.

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, we compute the riskfree return by rolling over the 3-month T-bill every quarter. For 1926–2001, the T-bill rates are taken from CRSP's Fama Risk Free Rates File. For our longer S&P 500 series, we augment this with US Commercial Paper Rates, New York City from Macaulay (1938).[6]

For the three CRSP series, we consider the sub-sample 1952–2001 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.[7] 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.

[5] Shiller's data is available at http://aida.econ.yale.edu/∼shiller/data.htm.

[6] Available at http://www.nber.org/databases/macrohistory/contents/chapter13.html.

[7] Available at http://www.stls.frb.org/fred/data/irates.html.

4.2 Empirical Results

4.2.1 Persistence of Predictor Variables

In Table 3, we report the 95% confidence interval of the autoregressive root ρ 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 yield spread $(y-r_1)$. The confidence interval is computed by applying the method of confidence belts (Stock (1991)) to the DF-GLS statistic. The autoregressive lag length $p \in [1, \overline{p}]$ for the predictor variable is estimated using BIC (Schwartz criterion). We set the maximum lag length $\overline{p}$ to 4 for annual, 8 for quarterly, and 12 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 $d-p$ 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 2001 (Panel A). The source of this difference can be explained by Figure 2, which is a time series plot of $d-p$ and $e-p$ at quarterly frequency. Around 1994, these valuation ratios begin to drift down to historical lows, making the processes look more like unit-root processes. The least persistent series is $y-r_1$, 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 standard normal — may be misleading. As shown in Section 3.3, whether conventional inference based on the t-test is reliable also depends on the correlation δ between the innovations to excess returns and the predictor variable. Hence, we report point estimates of δ in the fifth column of Table 3. As expected, the correlations for $d-p$ and $e-p$ 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 $\widehat{\delta}$ 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 3.3. For r_3 and $y-r_1$, $\widehat{\delta}$ 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.

4.2.2 Testing the Predictability of Returns

In this section, we construct valid confidence intervals for β to test the predictability of returns. Based on the power analysis in Section 3.5, our preferred test is the Bonferroni Q-test.

Our methodology and results can most easily be explained by the following graphical method, which can be implemented as a sequence of OLS regressions:

1. Run the OLS regressions (14) and (15), with the autoregressive lag length p estimated by BIC, to obtain $\widehat{\beta}$, $\widehat{\psi}_i$ $(i = 1, \ldots, p-1)$, and the standard error of $\widehat{\beta}$ which will be denoted by $\mathrm{SE}(\widehat{\beta})$. Using the OLS residuals, $\widehat{u}_t$ and $\widehat{v}_t$, compute $\widehat{\sigma}_u^2 = (T-2)^{-1}\sum_{t=1}^T \widehat{u}_t^2$, $\widehat{\sigma}_v^2 = (T-2)^{-1}\sum_{t=1}^T \widehat{v}_t^2$, $\widehat{\sigma}_{uv} = (T-2)^{-1}\sum_{t=1}^T \widehat{u}_t\widehat{v}_t$, and $\widehat{\delta} = \widehat{\sigma}_{uv}/(\widehat{\sigma}_u\widehat{\sigma}_v)$.

2. Construct a $100(1-\underline{\alpha}_1-\overline{\alpha}_1)\%$ confidence interval for ρ, denoted by $C_\rho(\alpha_1)$, using the appropriate values of $\underline{\alpha}_1$ and $\overline{\alpha}_1$ from Table 2 based on $\widehat{\delta}$.

3. For each value of ρ in $C_\rho(\alpha_1)$, compute an equal-tailed 90% confidence interval for β given ρ as follows. Regress $r_t - (\widehat{\sigma}_{uv}/\widehat{\sigma}_v^2)(x_t - \rho x_{t-1} - \sum_{i=1}^{p-1}\widehat{\psi}_i \Delta x_{t-i})$ onto a constant and x_{t-1}. Let $\widehat{\beta}(\rho)$ denote the coefficient on x_{t-1}. The confidence interval for β given ρ is $C_{\beta|\rho}(\alpha_2) = [\underline{\beta}(\rho,\alpha_2), \overline{\beta}(\rho,\alpha_2)]$, where

$$\begin{aligned}
\underline{\beta}(\rho,\alpha_2) &= \widehat{\beta}(\rho) - z_{\alpha_2/2}(1-\widehat{\delta}^2)^{1/2}\mathrm{SE}(\widehat{\beta}), \\
\overline{\beta}(\rho,\alpha_2) &= \widehat{\beta}(\rho) + z_{\alpha_2/2}(1-\widehat{\delta}^2)^{1/2}\mathrm{SE}(\widehat{\beta}).
\end{aligned}$$

4. Plot $C_{\beta|\rho}(\alpha_2)$ against ρ for all $\rho \in C_\rho(\alpha_1)$.

In practice, we only need to compute the confidence interval $C_{\beta|\rho}(\alpha_2)$ at the end points of $C_\rho(\alpha_1)$ since $\underline{\beta}(\rho,\alpha_2)$ and $\overline{\beta}(\rho,\alpha_2)$ are linear in ρ. Note that this results in a 10% two-sided test (i.e. 90% confidence interval) or a 5% one-sided test for predictability.

In reporting our confidence interval for β, we will scale it by $\widehat{\sigma}_v/\widehat{\sigma}_u$. In other words, we report the confidence interval for $\widetilde{\beta} = (\sigma_v/\sigma_u)\beta$ instead of β. Although this normalization does not affect inference, it is a more natural way to report results for two reasons. First, $\widetilde{\beta}$ has a natural interpretation as the coefficient in (14) when the errors in (14) and (15) are normalized to have unit variance. This is in the spirit of our statistical model (1)–(2), which assumed unit variance in the innovations. Second, by the equality

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

$\widetilde{\beta}$ 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 β to denote $\widetilde{\beta}$ throughout the rest of the paper.

In Figure 6, we plot the Bonferroni confidence interval for both $d-p$ and $e-p$ for annual and quarterly CRSP series (1927–2001). The solid lines represent the confidence interval based on the Bonferroni Q-test, and the dashed lines represent the confidence interval based on the Bonferroni t-test. The numerical procedure described in Section 3.4 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 confidence interval for ρ are chosen to result in a 5% one-sided test for β, uniformly in $c \in [-50, 5]$. Because of the asymmetry in the null distribution of the t-statistic, the confidence interval for ρ used for the right-tailed Bonferroni t-test differs from that used for 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 $d-p$ in the upper left panel, the Bonferroni confidence interval for β 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 Section 3.5. From Table 3, $\widehat{\delta} = -0.721$ and the confidence interval for c is $[-5.637, 4.097]$. 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 $d-p$ in the lower left panel of Figure 6, the evidence for predictability is weaker. In the relevant range of the confidence interval for ρ, the confidence interval for β 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 β lies above zero when $\rho \leq 0.988$. This means that if the true ρ 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 ρ, we cannot reject the null of no predictability.

In the upper right panel, we test for predictability in annual data using $e - p$ as the predictor variable. We find that stock returns are predictable with the Bonferroni Q-test but not with the Bonferroni t-test. In the lower right panel, we test predictability at the quarterly frequency using $e - p$ and obtain the same results. Again, the fact that the Bonferroni Q-test gives tighter confidence intervals can be explained by the power functions in Figure 5.

In Figure 7, we repeat the same exercise as Figure 6, using the quarterly CRSP data in the sub-sample 1952–2001. We report the plots for all four of our predictor variables: $d - p$ (upper left), $e - p$ (upper right), r_3 (lower left), and $y - r_1$ (lower right). For $d - p$, we find evidence for predictability if $\rho \leq 1.006$. 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 ρ, 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 ρ must be less than 0.998 before we can conclude that $e - p$ predicts returns. Taking account of the uncertainty in the true value of ρ, we cannot reject the null hypothesis $\beta = 0$.

The weaker evidence for predictability in the period since 1952 seems to be partly due to the fact that the valuation ratios appear more persistent when restricted to this sub-sample, so the confidence intervals for ρ contain rather large values of ρ that were excluded in Figure 6.

For r_3, the Bonferroni confidence interval for β lies strictly below zero for both the Q-test and the t-test over the entire confidence interval for ρ. For $y - r_1$, the evidence for predictability is similarly strong, with the confidence interval strictly above zero over the entire range of ρ. The power advantage of the Bonferroni Q-test over the Bonferroni t-test is small when δ 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 β using the t-test. In the sixth column, we report the Bonferroni confidence interval for the Q-test. In relation to Figures 6–7, we simply report the minimum and maximum values of β that the confidence bands achieve.

Focusing first on the full-sample results in Panel A, the Bonferroni Q-test rejects the null of no predictability for $e-p$ at all frequencies. For $d-p$, 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. Moving to the results for the sub-sample through 1994 in Panel B, we obtain qualitatively the same results using the Bonferroni Q-test with rejections for $e-p$ at all frequencies. Interestingly, the Bonferroni t-test gives similar results to the Bonferroni Q-test in this sub-sample. In this sub-sample, the power gains from using the Bonferroni Q-test appear to be minor.

In Panel C, we report the results for the sub-sample since 1952. In this sub-sample, we cannot reject the null hypothesis for $d-p$ or $e-p$. For the predictor variable r_3, we reject the null hypothesis except at annual frquency, and for $y-r_1$, we reject at all frequencies.

As we have seen in Figure 7, the weak evidence for predictability using the valuation ratios stems from the fact that the confidence intervals for ρ contain explosive values. If we could obtain tighter confidence intervals for ρ that exclude these values, the lower end of the confidence intervals for β would shrink, strengthening the evidence for predictability. In the last two columns of Table 4, we report how the lower end of the confidence interval for β changes if we impose the restriction $\rho \leq 1$. This corresponds to Lewellen's (2002) sup-bound Q-test that restricts the parameter space to $c \leq 0$. In terms of Figures 6–7, this is equivalent to discarding the region of the plots where $\rho > 1$. Note that under this restriction, the lower ends of the confidence intervals lie above zero for $d-p$ at all frequencies. So $d-p$ can predict returns in the sub-sample since 1952 if we can rule out explosive roots, consistent with Lewellen's results.

To summarize the empirical results, we find evidence for predictability with $e-p$, r_3, and $y-r_1$. The evidence for predictability using $d-p$ is much 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 greater power. The power gain is empirically relevant in the full sample through 2001.

4.3 Connection to Previous Empirical Results

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.

4.3.1 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 2 and sometimes greater than 3. Comparing the full sample through 2001 (Panel A) and the subsample through 1994 (Panel B), the evidence for predictability appears to have weakened in the last seven years. In the late 1990's, stock returns were high when $d - p$ and $e - p$ 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 7 observations to a sample of 114 (for S&P 500) or an addition of 28 data points to a sample of 272 (for quarterly CRSP). In fact, this sensitivity is evidence for the failure of first-order asymptotics. The t-statistic is not normally distributed under the null in finite samples, so the conventional critical values lead to wrong inference. Intuitively, when a predictor variable is persistent, its sample mean can change dramatically with an addition of a few data points. This is what happened in the late 1990's when valuation ratios reached historical lows. Since the t-statistic measures the covariance of excess returns and the valuation ratio, its value is sensitive to a shift in the sample mean. 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 that is robust to the persistence problem, we find that $e - p$

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 β in the fourth column of Table 4. As equations (26) and (27) show, the point estimate $\widehat{\beta}$ does not necessarily lie in the center of the robust confidence interval for β. Indeed, $\widehat{\beta}$ often falls toward the upper end of the Bonferroni confidence interval based on the Q-test for $d-p$ and $e-p$, and in a few cases, $\widehat{\beta}$ falls strictly above the confidence interval. This is a consequence of the upward finite sample bias of the OLS estimator due to the persistence of these predictor variables (cf. Stambaugh (1999) and Lewellen (2002)).

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 $d-p$ or $e-p$. As reported in Table 1, the unknown persistence parameter c must be less than about -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. Of course, one can also use the Q-test imposing the prior information $c < -70$, which leads to a more powerful test. Our empirical findings in Figures 6–7 confirm that there is strong evidence for predictability with $d-p$ or $e-p$ when $\rho \ll 1$. The difficulty is that the lower end 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 predictor variables r_3 and $y - r_1$, the correlation δ is sufficiently small that conventional inference based on the t-test leads to approximately the correct inference. This is confirmed in Panel C of Table 4 where the conventional t-test and the Bonferroni Q-test both reject the null.

4.3.2 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, $d - p$ has a smaller autoregressive coefficient than it does in monthly data and is less persistent in this sense. Over periods of several years, $d - p$ has an even lower autoregressive coefficient. 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 (2002) and Torous, Valkanov, and Yan (2001) for theoretical 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 (2002) 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, which long-horizon regression tests do not.

4.3.3 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 (2002) 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 α-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 $\delta < 0$ and we know *a priori* that $\rho \leq 1$. As we have shown in Panel C of Table 4, the 5% one-sided test using monthly $d - p$ rejects when $\rho = 1$, confirming Lewellen's empirical findings.

Although Lewellen's assumption that $\rho \leq 1$ may initially seem reasonable based on finance theory, we may not want to impose such a strong parametric assumption on the data. Even if we were to know with certainty that the dividend-price ratio is stationary, we do not know that the log dividend-price ratio is an AR(1) with $\rho \leq 1$. If the true data generating process for the dividend-price ratio is stationary but with some nonlinear effects, an AR(1) with ρ slightly greater than one may be a better approximation to the true process than an AR(1) with $\rho \leq 1$. For this reason, we have considered a flexible parametric model (15) for the predictor variable, allowing for the possibility that $\rho > 1$. 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) (i.e. $\psi_i = 0$ $(i = 1, \ldots, p - 1)$ in (15)).

Torous, Valkanov, and Yan (2001) develop a test of predictability that is conceptually similar to ours, constructing Bonferroni confidence intervals for β. One difference from our approach is that they construct the confidence interval for ρ using the Dickey-Fuller t-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 β given ρ. Their choice of the long-horizon t-test was motivated by their objective of highlighting the pitfalls of long-horizon regressions.

One key insight of Torous, Valkanov, and Yan (2001) is that the evidence for the predictability of returns with these persistent variables depends critically on the unknown degree of persistence. Because we cannot estimate the degree of persistence consistently, the evidence for predictability can be ambiguous. This point is illustrated in Figures 6–7, where we find that $d - p$ predicts returns if its autoregressive root ρ is sufficiently small. In this paper, we have confirmed their finding that the evidence for predictability by $d - p$ 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, the return process should have a unit or 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.[8] Lanne's empirical finding agrees with ours and those of Torous, Valkanov, and Yan (2001). From Figures 6–7, we see that if in fact $\beta > 0$, the degree of persistence in $d - p$ or $e - p$ must be sufficiently small. In addition, we find evidence for predictability using $y - r_1$ in the post-1952 sample, 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 $y - r_1$.

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.

[8] 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.

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 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 innovations to stock returns. Using our pretest, we find that the conventional t-test can lead to misleading inferences for these 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 weakly predicts returns only at an annual frequency. In the post-1952 sample, there is less evidence for predictability, 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

Throughout this appendix, we assume that Assumption 1 holds, $c = T(\rho - 1)$ is fixed, and $b = T(\beta - \beta_0)$ is fixed. Collecting results from Phillips (1987, Lemma 1), Chan and Wei (1988, Theorem 2.4), and Cavanagh, Elliott, and Stock (1995), we have the following convenient lemma.

Lemma 1 *Let $\eta_t = (u_t - (\sigma_{uv}/\sigma_v^2)v_t)/(\sigma_u(1-\delta^2)^{1/2})$. The following limits hold jointly:*

1. $T^{-2}\sum_{t=1}^T x_{t-1}^{\mu 2} \Rightarrow \omega^2 \int J_c^\mu(s)^2 ds$,
2. $T^{-1}\sum_{t=1}^T x_{t-1}^{\mu} v_t \Rightarrow \sigma_v \omega \int J_c^\mu(s) dW(s)$,
3. $T^{-1}\sum_{t=1}^T x_{t-1}^{\mu} \eta_t \Rightarrow \omega (\int J_c^\mu(s)^2 ds)^{1/2} Z$.

To obtain the asymptotic distributions of $N(\beta_0, \bar{b})$, $P(\beta_0, \bar{b}, \rho)$, and $Q(\beta_0, \rho)$, we first note that these statistics can be written as

$$\begin{aligned}
N(\beta_0, \bar{b}) &= (2\bar{b}b - \bar{b}^2)T^{-2}\sum_{t=1}^T x_{t-1}^{\mu 2} + 2\bar{b}\left[\frac{\sigma_{uv}}{\sigma_v^2}T^{-1}\sum_{t=1}^T x_{t-1}^{\mu} v_t + \sigma_u(1-\delta^2)^{1/2}T^{-1}\sum_{t=1}^T x_{t-1}^{\mu}\eta_t\right], \\
P(\beta_0, \bar{b}, \rho) &= (2\bar{b}b - \bar{b}^2)T^{-2}\sum_{t=1}^T x_{t-1}^{\mu 2} + 2\bar{b}\sigma_u(1-\delta^2)^{1/2}T^{-1}\sum_{t=1}^T x_{t-1}^{\mu}\eta_t, \\
Q(\beta_0, \rho) &= \frac{b(T^{-2}\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}{\sigma_u(1-\delta^2)^{1/2}} + \frac{T^{-1}\sum_{t=1}^T x_{t-1}^{\mu}\eta_t}{(T^{-2}\sum_{t=1}^T x_{t-1}^{\mu 2})^{1/2}}.
\end{aligned}$$

Then an application of Lemma 1 results in (18), (21), and (22), respectively.

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.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock, 1995, Inference in models with nearly integrated regressors, *Econometric Theory* 11, 1131–1147.

Chan, N. H., and C. Z. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes, *Annals of Statistics* 16, 367–401.

Cochrane, John H., 1999, New facts in finance, NBER Working Paper 7169.

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.

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.

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, 2002, Predicting returns with financial ratios, Working paper, MIT Sloan School of Management.

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).

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.

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, Working paper, The Anderson School, UCLA.

Valkanov, Rossen, 2002, Long-horizon regressions: Theoretical results and applications, *Journal of Financial Economics*, forthcoming.

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 being considered is $\beta = \beta_0$ against the alternative $\beta > \beta_0$. For a given δ, the size of the t-test is greater than 7.5% if $c \in (\underline{c}, \overline{c})$. Size is less less than 7.5% for all c if $\delta \leq -0.125$.

δ	$\underline{c}$	$\overline{c}$	δ	$\underline{c}$	$\overline{c}$
-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 DF-GLS Confidence Interval for Bonferroni Q-test

This table reports the significance level of the confidence interval for the largest autoregressive root ρ, computed by inverting the DF-GLS test, that sets the size of the one-sided Bonferroni Q-test to 5%. Using the notation (28), the confidence interval $C_\rho(\alpha_1) = [\underline{\rho}(\underline{\alpha}_1), \overline{\rho}(\overline{\alpha}_1)]$ for ρ results in a 90% Bonferroni confidence interval $C_\beta(0.1)$ for β when $\alpha_2 = 0.1$.

δ	$\underline{\alpha}_1$	$\overline{\alpha}_1$	δ	$\underline{\alpha}_1$	$\overline{\alpha}_1$
-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 Model Parameters

This table reports estimates of parameters for the predictive regression model. The data series are annual S&P 500 and CRSP at annual, quarterly, and monthly frequencies. The predictor variables are log dividend-price ratio ($d-p$), log earnings-price ratio ($e-p$), 3-month T-bill rate (r_3), and long-short yield spread ($y-r_1$). p is the estimated autoregressive lag length for the predictor variable, and $\widehat{\delta}$ is the estimated correlation between innovations to excess stock returns and the predictor variable. The last two columns are 95% confidence intervals for the largest autoregressive root (ρ) 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	$\widehat{\delta}$	DF-GLS	95% CI: ρ	95% CI: c
			Panel A: Full Sample				
S&P 500	1881–2001	$d-p$	3	-0.843	-1.058	[0.937,1.031]	[-7.483,3.746]
	(122)	$e-p$	1	-0.963	-2.599	[0.801,0.982]	[-24.093,-2.235]
Annual	1927–2001	$d-p$	1	-0.721	-0.773	[0.925,1.055]	[-5.637,4.097]
	(76)	$e-p$	1	-0.960	-1.951	[0.789,1.017]	[-15.798,1.275]
Quarterly	1927–2001	$d-p$	1	-0.942	-1.448	[0.964,1.010]	[-10.651,3.009]
	(301)	$e-p$	1	-0.986	-1.918	[0.949,1.005]	[-15.434,1.417]
Monthly	1927–2001	$d-p$	2	-0.950	-1.450	[0.988,1.003]	[-10.672,3.003]
	(901)	$e-p$	2	-0.982	-1.756	[0.985,1.002]	[-13.701,1.998]

(continued on next page)

Series	Sample (Obs)	Variable	p	$\widehat{\delta}$	DF-GLS	95% CI: ρ	95% CI: c
			Panel B: Sample through 1994				
S&P 500	1881–1994	$d-p$	3	-0.841	-2.860	[0.752,0.964]	[-27.831,-4.082]
	(115)	$e-p$	1	-0.959	-3.492	[0.667,0.916]	[-37.979,-9.521]
Annual	1927–1994	$d-p$	1	-0.695	-2.081	[0.745,1.010]	[-17.348,0.687]
	(69)	$e-p$	1	-0.962	-2.849	[0.593,0.941]	[-27.658,-4.022]
Quarterly	1927–1994	$d-p$	1	-0.941	-2.635	[0.910,0.991]	[-24.585,-2.473]
	(273)	$e-p$	1	-0.988	-2.811	[0.900,0.986]	[-27.073,-3.701]
Monthly	1927–1994	$d-p$	2	-0.948	-2.592	[0.971,0.997]	[-24.006,-2.195]
	(817)	$e-p$	2	-0.983	-2.655	[0.970,0.997]	[-24.852,-2.577]
			Panel C: Sample from 1952				
Annual	1953–2001	$d-p$	1	-0.744	-0.068	[0.944,1.093]	[-2.725,4.556]
	(50)	$e-p$	1	-0.957	-1.172	[0.831,1.073]	[-8.305,3.575]
		r_3	1	-0.037	-1.812	[0.708,1.038]	[-14.288,1.853]
		$y-r_1$	1	-0.202	-3.242	[0.311,0.853]	[-33.746,-7.189]
Quarterly	1952–2001	$d-p$	1	-0.977	0.056	[0.988,1.023]	[-2.353,4.612]
	(200)	$e-p$	1	-0.980	-0.783	[0.971,1.021]	[-5.687,4.088]
		r_3	4	-0.120	-1.462	[0.945,1.015]	[-10.783,2.972]
		$y-r_1$	2	-0.079	-3.209	[0.832,0.965]	[-33.237,-6.934]
Monthly	1952–2001	$d-p$	1	-0.966	0.154	[0.996,1.008]	[-2.105,4.657]
	(600)	$e-p$	1	-0.982	-0.531	[0.993,1.007]	[-4.418,4.298]
		r_3	2	-0.084	-1.550	[0.981,1.005]	[-11.625,2.733]
		$y-r_1$	1	-0.055	-4.523	[0.904,0.964]	[-57.580,-21.384]

Table 4: Test of Predictability

This table reports statistics used to infer the predictability of excess stock returns. The data series are annual S&P 500 and CRSP at annual, quarterly, and monthly frequencies. (See Table 3 for the sample periods and the number of observations.) The predictor variables are log dividend-price ratio ($d-p$), log earnings-price ratio ($e-p$), 3-month T-bill rate (r_3), and long-short yield spread ($y-r_1$). In the third and fourth columns, the table reports the t-statistic and point estimate $\widehat{\beta}$ from an OLS regression of returns onto the predictor variable. The next two columns report the 90% Bonferroni confidence intervals for β using the t-test and Q-test, respectively. The final column reports the lower end of the Bonferroni confidence interval when the constraint $\rho \leq 1$ on the largest autoregressive root of the predictor variable is imposed.

Series	Variable	t-stat	$\widehat{\beta}$	90% CI: β		Low CI β
				t-stat	Q-stat	($\rho \leq 1$)
			Panel A: Full Sample			
S&P 500	$d-p$	1.486	0.076	[-0.069,0.124]	[-0.029,0.114]	-0.021
	$e-p$	2.363	0.113	[-0.004,0.170]	[0.013,0.188]	—
Annual	$d-p$	2.169	0.112	[-0.026,0.165]	[0.000,0.172]	0.012
	$e-p$	2.490	0.154	[-0.015,0.223]	[0.022,0.238]	—
Quarterly	$d-p$	1.754	0.030	[-0.020,0.046]	[-0.013,0.036]	-0.011
	$e-p$	2.710	0.046	[-0.002,0.064]	[0.006,0.057]	—
Monthly	$d-p$	1.383	0.007	[-0.008,0.012]	[-0.007,0.009]	-0.006
	$e-p$	2.472	0.013	[-0.002,0.018]	[0.000,0.016]	—

(continued on next page)

Series	Variable	t-stat	$\widehat{\beta}$	90% CI: β		Low CI β
				t-stat	Q-stat	($\rho \leq 1$)
			Panel B: Sample through 1994			
S&P 500	$d-p$	2.124	0.142	[-0.007,0.228]	[-0.040,0.197]	—
	$e-p$	3.238	0.191	[0.064,0.268]	[0.088,0.318]	—
Annual	$d-p$	2.960	0.210	[0.038,0.299]	[0.054,0.329]	—
	$e-p$	3.434	0.282	[0.093,0.385]	[0.128,0.446]	—
Quarterly	$d-p$	2.283	0.053	[-0.002,0.080]	[-0.007,0.075]	—
	$e-p$	3.576	0.082	[0.027,0.110]	[0.027,0.109]	—
Monthly	$d-p$	1.798	0.013	[-0.004,0.021]	[-0.006,0.020]	—
	$e-p$	3.284	0.023	[0.006,0.031]	[0.007,0.031]	—
			Panel C: Sample from 1952			
Annual	$d-p$	1.876	0.108	[-0.048,0.170]	[-0.017,0.169]	0.015
	$e-p$	1.480	0.098	[-0.095,0.157]	[-0.049,0.183]	-0.019
	r_3	-1.405	-0.119	[-0.262,0.019]	[-0.262,0.020]	—
	$y-r_1$	1.729	0.218	[-0.008,0.416]	[0.002,0.448]	—
Quarterly	$d-p$	1.872	0.031	[-0.018,0.044]	[-0.010,0.026]	0.006
	$e-p$	1.562	0.025	[-0.022,0.036]	[-0.014,0.032]	-0.001
	r_3	-1.981	-0.048	[-0.093,-0.010]	[-0.093,-0.011]	—
	$y-r_1$	2.598	0.123	[0.041,0.199]	[0.038,0.196]	—
Monthly	$d-p$	1.846	0.010	[-0.006,0.014]	[-0.004,0.008]	0.001
	$e-p$	1.502	0.008	[-0.007,0.012]	[-0.005,0.009]	0.000
	r_3	-2.732	-0.020	[-0.033,-0.008]	[-0.033,-0.008]	—
	$y-r_1$	3.615	0.059	[0.032,0.086]	[0.031,0.086]	—

Figure 1. Local Asymptotic Power under First-Order Asymptotics. This figure plots the power of the one-sided Q-test and t-test when the predictor variable is an AR(1). The local alternatives being considered are $b = \sqrt{T}(\beta_1 - \beta_0)$. $\rho = 0.99, 0.75$ is the autoregressive root of the predictor variable, and $\delta = -0.95, -0.75$ is the correlation between innovations to returns and the predictor variable.

Figure 2. Time Series Plot of Valuation Ratios. This figure plots the log dividend-price ratio $(d - p)$ for the CRSP NYSE/AMEX portfolio and the log earnings-price ratio $(e - p)$ for the S&P 500 portfolio at quarterly frequency. Earnings are smoothed by taking a ten year moving average.

Figure 3. Local Asymptotic Power under Local-to-Unity Asymptotics. This figure plots the power of the one-sided Q-test and t-test when the predictor variable contains a local-to-unit root. It also plots the power envelope for point optimal tests. The local alternatives being considered are $b = T(\beta_1 - \beta_0) > 0$. $c = -2, -20$ is the local-to-unity parameter, and $\delta = -0.95, -0.75$ is the correlation between innovations to returns and the predictor variable.

Figure 4. 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 predictor variable has an autoregressive root that is local-to-unity. The null hypothesis is $\beta = \beta_0$ against the one-sided alternative $\beta > \beta_0$. c is the local-to-unity parameter, and δ is the correlation between innovations to returns and the predictor variable.

Figure 5. Local Asymptotic Power of Feasible Tests. This figure plots the power of two Bonferroni tests (based on the Q-test and the t-test), the sup-bound Q-test, and the infeasible t-test that assumes knowledge of the local-to-unity parameter. The local alternatives being considered are $b = T(\beta_1 - \beta_0) > 0$. $c = -2, -20$ is the local-to-unity parameter, and $\delta = -0.95, -0.75$ is the correlation between innovations to returns and the predictor variable.

Figure 6. Bonferroni Confidence Interval (Annual and Quarterly, 1927–2001). This figure plots the 90% confidence interval for β over the confidence interval for ρ. The significance level for ρ is chosen to result in a 90% Bonferroni confidence interval for β. The solid (dashed) line is the confidence interval for β computed by inverting the Q-test (t-test).

The data series used are annual and quarterly CRSP (1927–2001). The predictor variables are log dividend-price ratio and log earnings-price ratio.

Figure 7. Bonferroni Confidence Interval (Quarterly, 1952–2001). This figure plots the 90% confidence interval for β over the confidence interval for ρ. The significance level for ρ is chosen to result in a 90% Bonferroni confidence interval for β. The solid (dashed) line is the confidence interval for β computed by inverting the Q-test (t-test). The data series used is quarterly CRSP (1952–2001). The predictor variables are log dividend-price ratio, log earnings-price ratio, 3-month T-bill rate, and long-short yield spread.

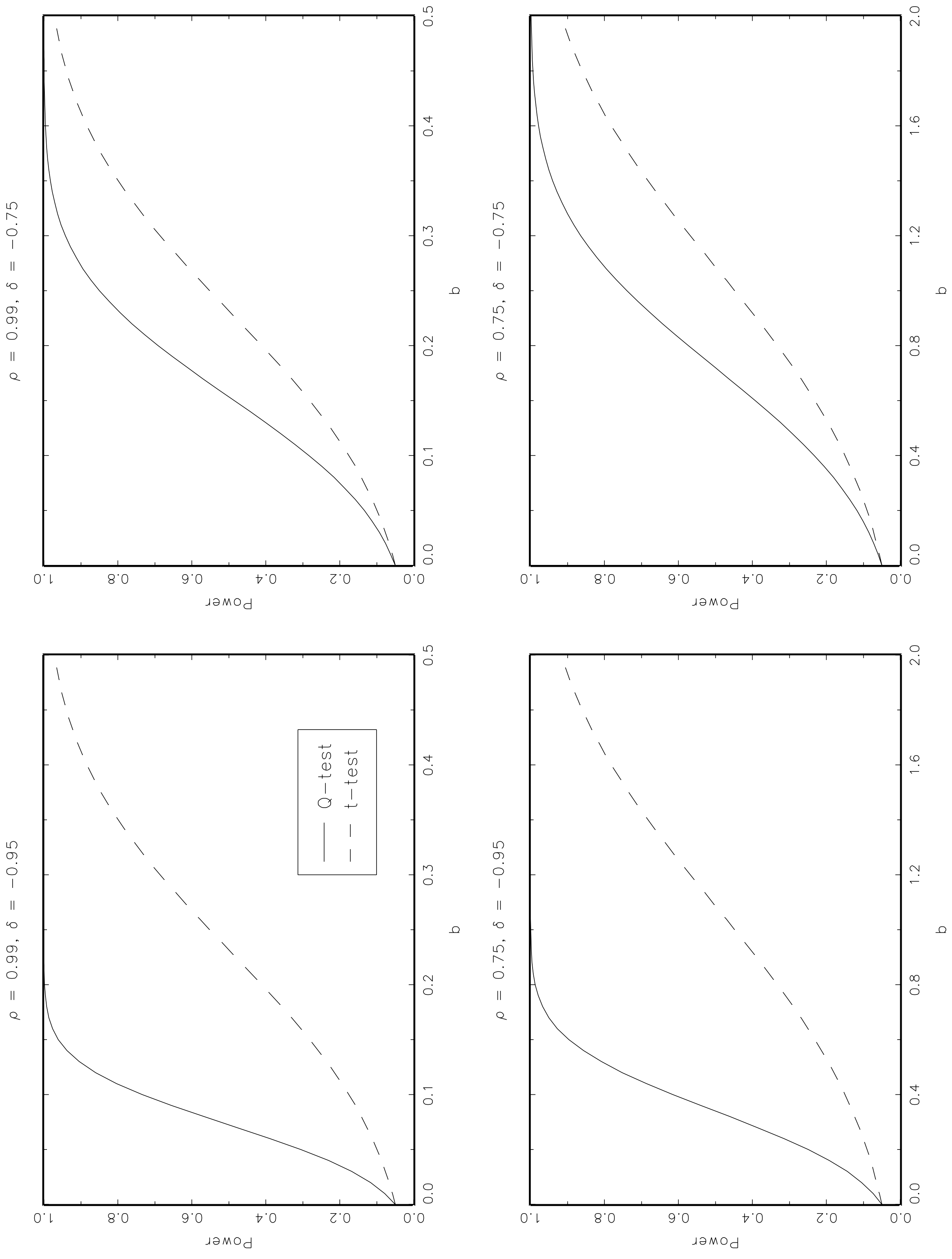

Figure 1: Local Asymptotic Power under First-Order Asymptotics

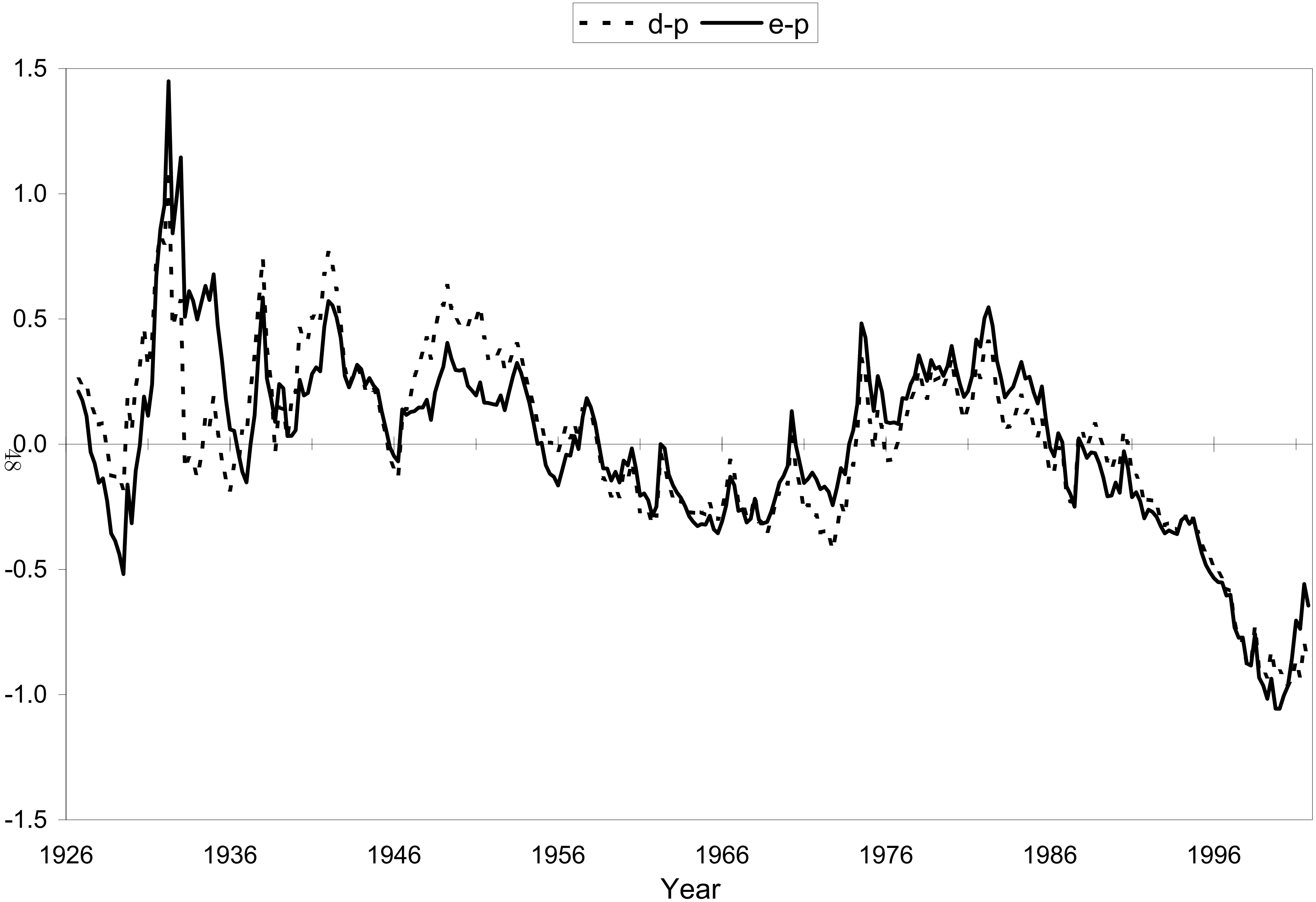

Figure 2: Time Series Plot of Valuation Ratios

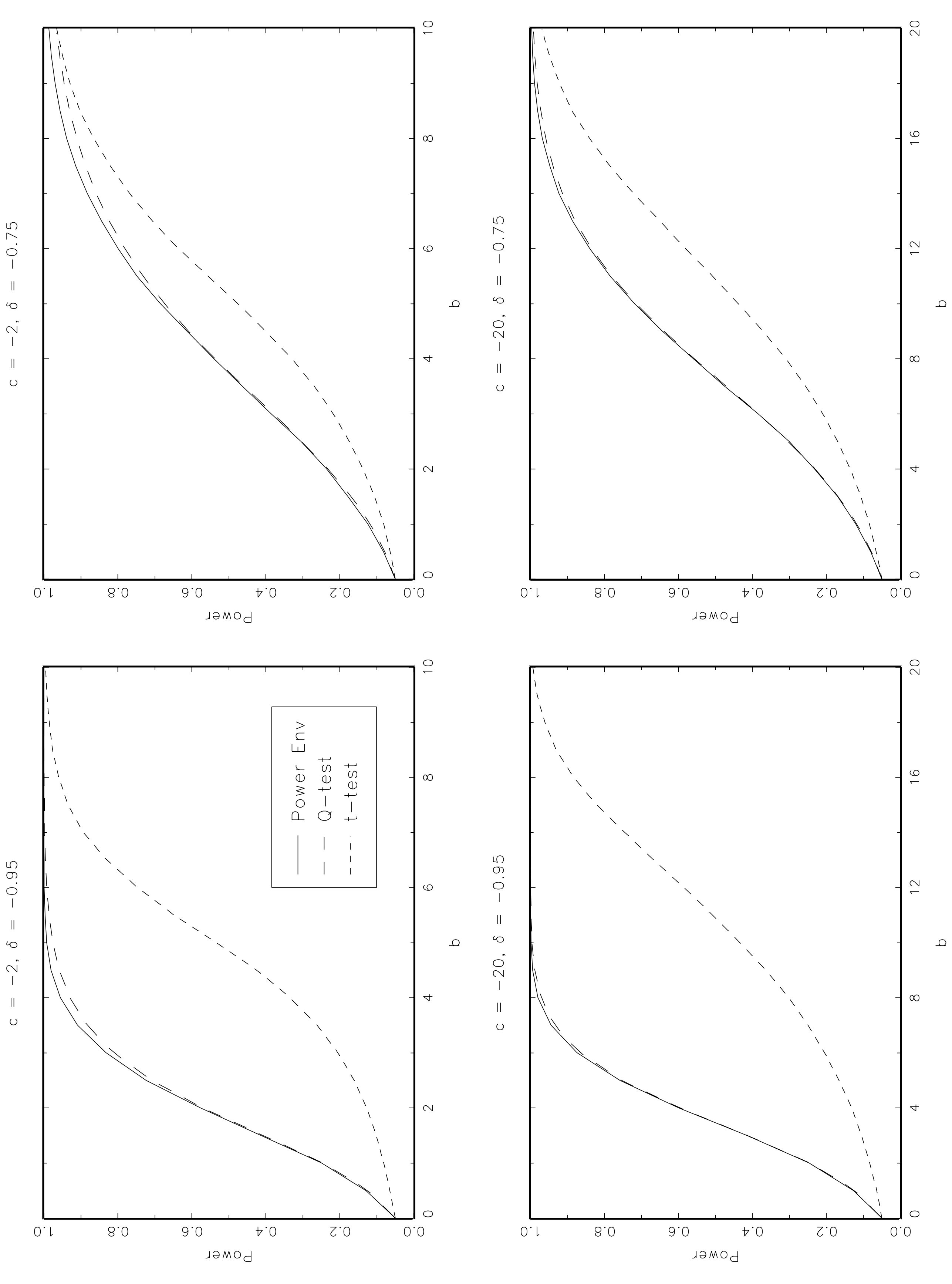

Figure 3: Local Asymptotic Power under Local-to-Unity Asymptotics

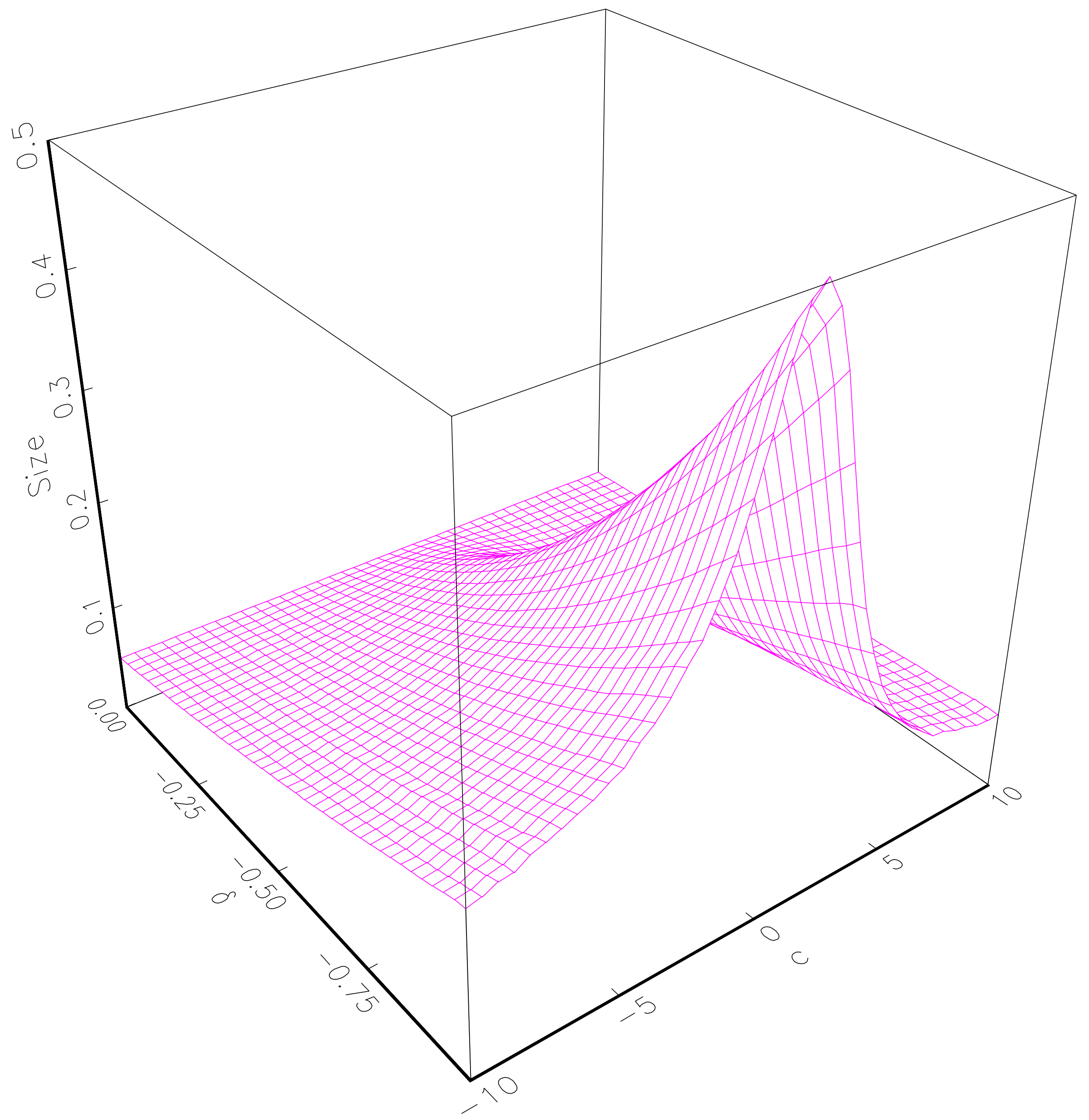

Figure 4: Asymptotic Size of One-Sided t-test at 5% Significance

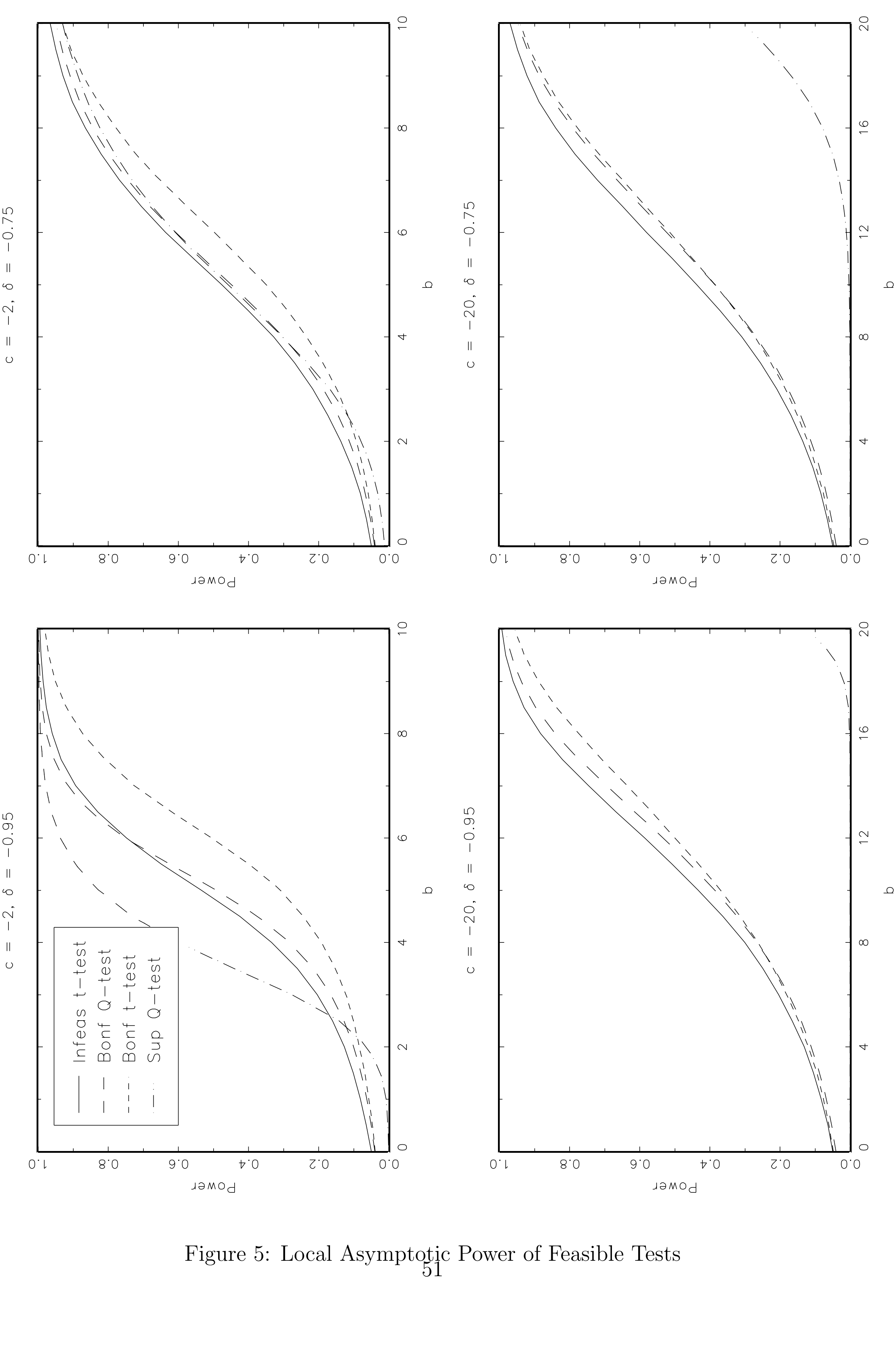

Figure 5: Local Asymptotic Power of Feasible Tests

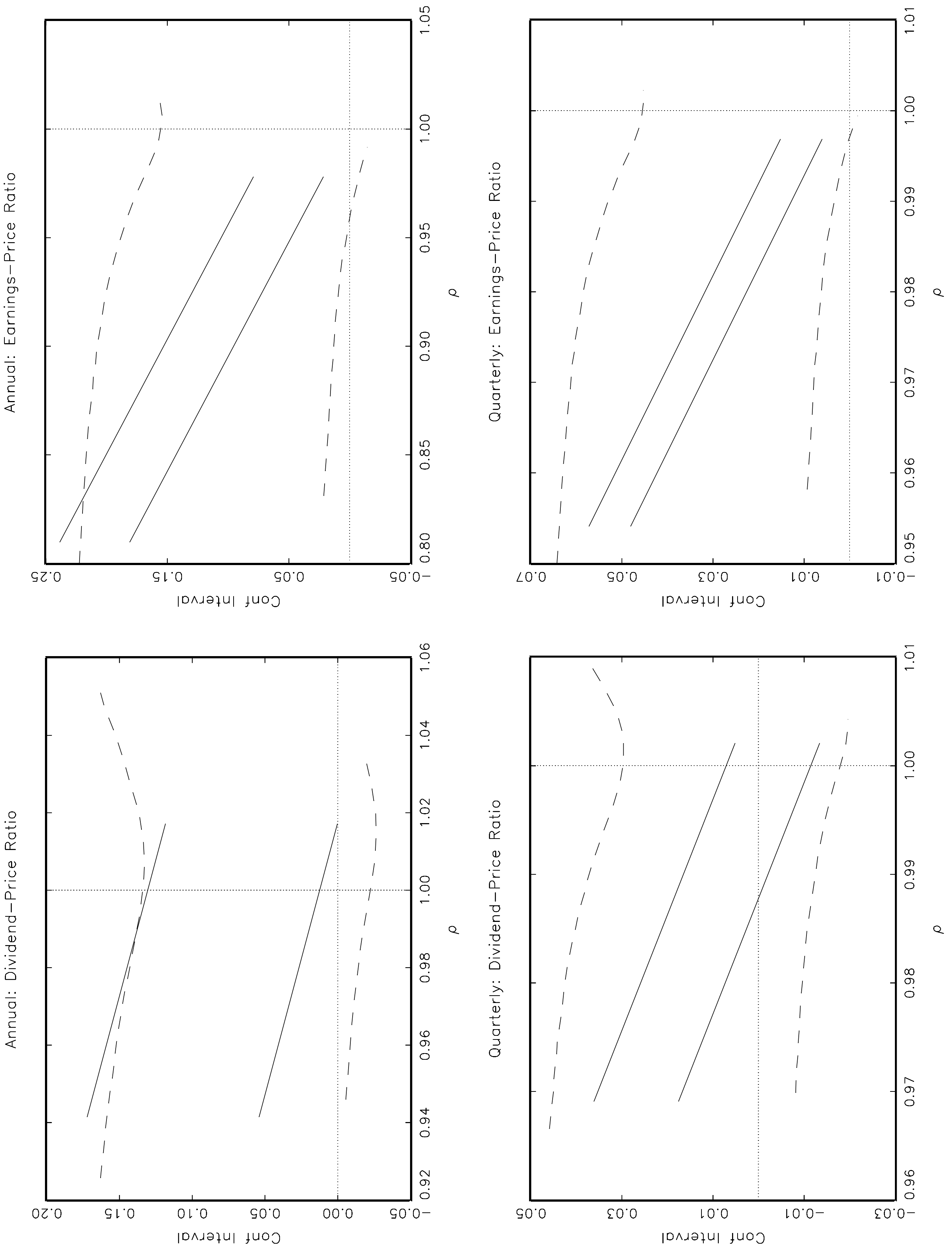

Figure 6: Bonferroni Confidence Interval (Annual and Quarterly, 1927–2001)

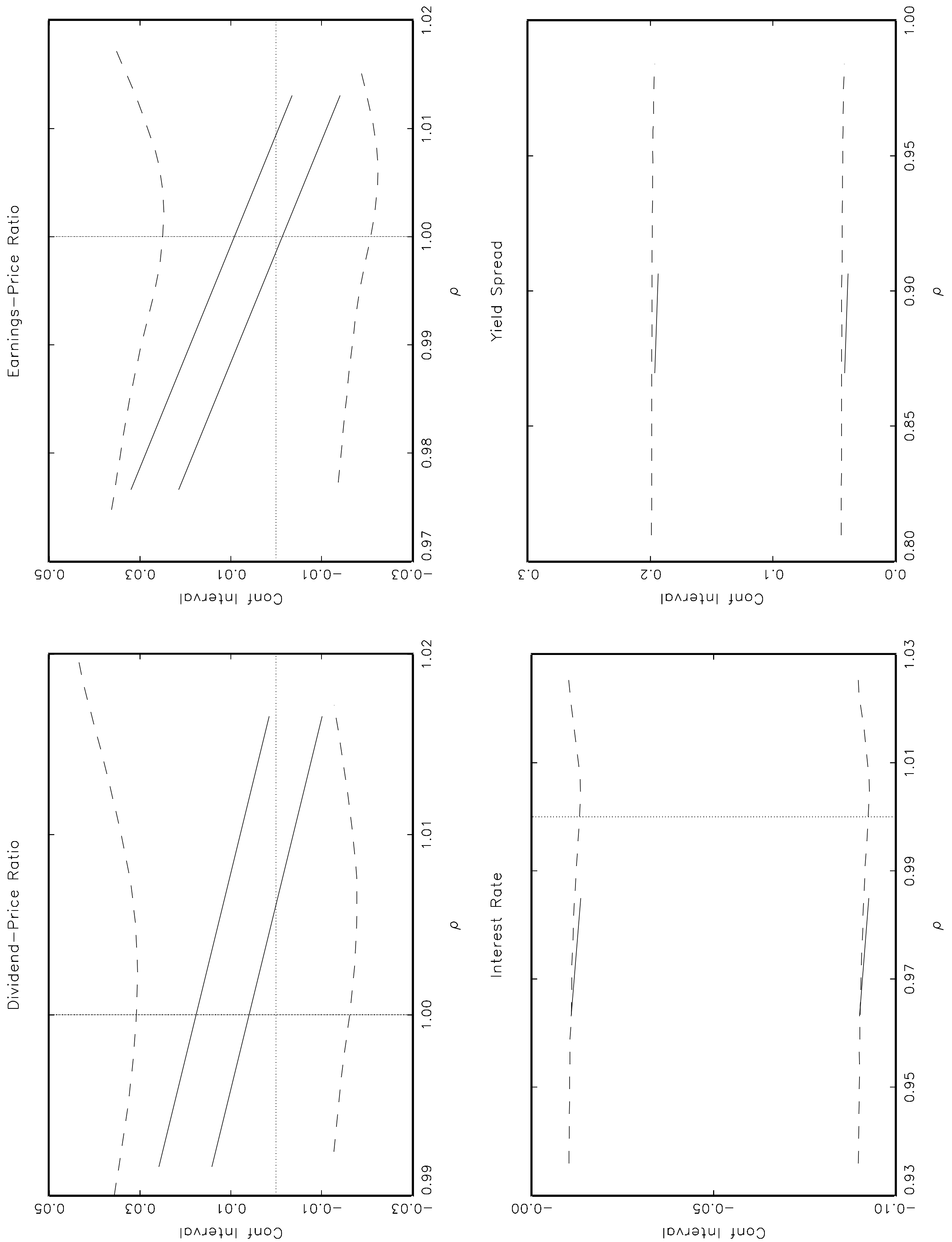

Figure 7: Bonferroni Confidence Interval (Quarterly, 1952–2001)