Boundaries of Predictability: Noisy Predictive Regressions*

Walter Torous[a]
The Anderson School
University of California, Los Angeles

Rossen Valkanov[b]
The Anderson School
University of California, Los Angeles

December 2000

* We thank Yacine Ait-Sahalia, John Cochrane, Pedro Santa-Clara, Mark Watson, and Ivo Welch for useful discussions and comments. All remaining errors are our own.

[a] The Anderson School; University of California, Los Angeles; 110 Westwood Plaza; Los Angeles, CA 90095-1481; Phone: (310) 206-6077; Fax: (310) 206-5455; Email: walter.n.torous@anderson.ucla.edu.

[b] Corresponding author. The Anderson School; University of California, Los Angeles; 110 Westwood Plaza; Los Angeles, CA 90095-1481; Phone: (310) 825-7246; Fax: (310) 206-5455; Email: rossen.valkanov@anderson.ucla.edu.

Boundaries of Predictability: Noisy Predictive Regressions

December 2000

Abstract

Even if returns are truly forecasted by variables such as the dividend yield, the noise in such a predictive regression may overwhelm the signal of the conditioning variable and render estimation, inference and forecasting unreliable. Unfortunately, traditional asymptotic approximations are not suitable to investigate the small sample properties of forecasting regressions with excessive noise. To systematically analyze predictive regressions, it is useful to quantify a forecasting variable's signal relative to the noisiness of returns in a given sample. We define an index of signal strength, or information accumulation, by renormalizing the signal-noise ratio. The novelty of our parameterization is that this index explicitly influences rates of convergence and can lead to inconsistent estimation and testing, unreliable R^2s, and no out-of-sample forecasting power. Indeed, we prove that if the signal-noise ratio is close to zero, as is the case for many of the explanatory variables previously suggested in the finance literature, model based forecasts will do no better than the corresponding simple unconditional mean return. Our analytic framework is general enough to capture most of the previous findings surrounding predictive regressions using dividend yields and other persistent forecasting variables.

Boundaries of Predictability: Noisy Predictive Regressions

In the presence of persistent time-varying expected returns it is well known that realized returns can be extremely variable even if expected returns themselves are not. Because of this noise, it is difficult to rely on past returns to statistically distinguish the random walk model from alternative models characterized by slowly mean reverting stock prices.[1] Consequently, researchers have turned their attention to regressions using forecasting variables other than past returns to predict stock returns. These forecasting variables include, among others, the dividend yield, term and credit spreads as well as short term and long term rates of interest.

This paper argues that the noise in realized returns which plagues empirical tests of the random walk model using lagged returns also adversely affects predictive regressions which use conditioning variables such as the dividend yield. Dividend yields are extremely persistent but, in contrast to returns, are not very variable. Even if there are sound economic arguments suggesting that dividend yields do contain valuable information about future stock returns, for example, by appealing to Campbell and Shiller's (1988) dynamic Gordon growth model, there is so much noise in stock returns and concomitantly so little variability in dividend yields (and all other frequently used predictors) that reliable estimation, inference, and forecasting cannot be carried out. While the adverse role of noise in statistical estimation may not come as a surprise, this paper provides a systematic analysis of its full implications for predictive regression models.

Unfortunately, we cannot rely upon traditional asymptotic theory to investigate the finite sample properties of noisy predictive regressions. In traditional theory as the sample size increases, the underlying signal is amplified by construction, so that as a result the least squares estimator becomes arbitrarily close to the true parameter value.[2] In other words, as the sample size increases, the signal-noise ratio is not held constant but rather increases at the same rate. The appropriate analysis of noise in predictive regressions, however, requires an asymptotic theory in which the signal-noise ratio does not necessarily grow with the sample size. For example, enlarging the sample size by decreasing the sampling interval for a fixed time span, say, by going from monthly to daily observations, may very well *increase* noise more so than information.[3] In general, for the

[1] See, among others, Shiller and Perron (1985), Summers (1986), and Poterba and Summers (1988). Cochrane (1988) also explores this issue in the context of estimating the dynamics of GNP.

[2] This is a loose definition of consistency.

[3] This is precisely the point raised by Shiller and Perron (1985) and, in particular, Perron (1989).

sample sizes typically encountered in regressions involving forecasts of returns, it is not appropriate to rely on a theory which at least asymptotically posits that noise and its effects are arbitrarily small.

By contrast, we incorporate noise by parameterizing the signal-noise ratio in predictive regressions to be a function of the sample size and explicitly control the rate of information (or signal) accumulation with the increasing sample size. The novelty of our parameterization is that the rate of information accumulation explicitly influences the rate of convergence and can lead to inconsistent estimation and testing in predictive regressions as well as unreliable R^2s. Such a parameterization allows us to analytically investigate the small sample properties of forecasting relations when the signal-noise ratio is small.

Using this framework, we provide an explicit measure or index of the informativeness of a posited explanatory variable relative to the noise present in returns. We demonstrate that this index governs the small sample behavior of predictive regressions and, furthermore, provides a boundary delineating how much forecasting power can be expected from a predictor. For example, we find that the index of the informativeness of the relative short term rate of interest exceeds that of other explanatory variables including the dividend yield but, in general, none of the predictors examined in this paper are particularly informative when measured against the noise present in returns. It is not surprising then that Bossaerts and Hillion (1999) and Goyal and Welch (1999) conclude that the ability of the dividend yield and other variables to forecast stock returns out-of-sample is abysmal. In fact, we can use our asymptotic approximation to show that even if the conditioning variables are informative about future stock returns, with so much noise, the forecasts produced with this correctly specified model will not do better in a mean squared error sense than the unconditional mean of past stock returns.

The adverse effects of noise analyzed in this paper are unrelated to the issues surrounding forecasting regressions raised by Stambaugh (1999). Stambaugh demonstrates that if the disturbances of the predictive regression and the autoregression describing the conditioning variable's dynamics are contemporaneously correlated then the presence of a predetermined variable results in a small sample bias in the predictive regression's slope coefficient. While this bias disappears asymptotically, our results do not and, in fact, continue to hold even if these disturbances are contemporaneously uncorrelated. Our conclusions also do not rely upon spurious regressions arguments (Granger and

Newbold (1974)), a concern that has been raised since at least Goetzmann and Jorion (1993). A spurious regression arises if unrelated integrated variables are regressed against one another. We, by contrast, assume returns and a forecasting variable, like the dividend yield, to be related. As noted by Ferson, Sarkissian, and Simin (2000), the traditional asymptotic theory applied to spurious regressions (Phillips (1986)) does not adequately describe predictive regressions in financial economics because of the sizable noise component in realized returns which is assumed absent in the traditional asymptotic theory. This, however, is precisely the issue addressed by our asymptotic theory and we establish that in the presence of this noise a spurious relation between returns and persistent forecasting variables is unlikely. To our knowledge, this is the first paper to offer a systematic and analytic study of predictive regressions with small signal-noise ratio.

The plan of this paper is as follows. In Section 1, assuming that a posited explanatory variable is informative about stock returns, we present Monte Carlo evidence investigating the properties of the resulting one-period ahead predictive regression as more noise is introduced into the estimation. We document that reliable estimation and inference cannot be carried out in the presence of signal-noise ratios typically encountered in practice. Section 2 provides an asymptotic theory to explain the properties of predictive regressions in the presence of varying signal-noise ratios. We argue that this asymptotic theory provides a better guide to understanding the results of predictive regressions typically found in the literature where the predictors are persistent but not very variable. Section 3 investigates the role of noise in both in-sample fit as well as out-of-sample forecasting. Estimators of the signal-noise ratio index are proposed in Section 4 and are used to assess the informativeness of the dividend yield, term and credit spreads and short term and long term rates of interest. Our conclusions are presented in Section 5.

1 Noisy Predictive Regressions

We consider a standard predictive regression in which an explanatory variable, X_t, is used to predict excess stock returns, r_t:

$$r_{t+1} = \mu + \beta X_t + \varepsilon_{t+1} \tag{1}$$

$$X_{t+1} = \mu_x + \phi X_t + u_{t+1} \tag{2}$$

where ε_t and u_t are random disturbances with mean zero, variances σ_ε^2 and σ_u^2, respectively, and covariance $\sigma_{\varepsilon u}$.

Two distinct issues are addressed in this paper. First, returns are observed to be extremely volatile when compared to many predictive variables. This, of course, is nothing but a restatement of the observation that the coefficient of determination, R^2, in the regression (1) is typically found to be quite low when many of the predictors suggested in the finance literature are used as explanatory variables. Unfortunately, the fact that only a small signal is present in the data has largely been ignored by appealing to traditional asymptotic arguments. According to these arguments, a small signal in the data is not a concern in a standard regression setting because as the number of observations T increases, the signal is assumed to increase at the same rate.[4] In practice, however, we deal with samples of fixed length and so we cannot increase the number of observations in order to increase the precision of our estimates. One might argue that increasing the frequency of observations, say from monthly to daily observations, rather than lengthening the sample period, may provide us with more information and, consequently, more precise estimates. But this need not be the case as increasing the frequency of observations can actually increase noise and so further obscure the small signal in the data.

Second, most explanatory variables are very persistent in the sense that the autoregressive coefficient ϕ in (2) is close to one and unit-root tests have difficulty in rejecting the null hypothesis that these variables are integrated. The fact that a nearly integrated explanatory variable is used to forecast stationary returns raises a number of econometric issues. In fact, this estimation can only make sense if, as pointed out earlier, the persistent variable's signal is small when compared to the stationary noise component arising from unexplained fluctuations in returns. Similar problems have been addressed by, among others, Poterba and Summers (1988), Cochrane (1988) and, most recently, Ferson Sarkissian, and Simin (2000).

It is worth emphasizing that if the system (1)-(2) is used to forecast returns then the persistence of the predictor is intricately linked with the overwhelming noise in returns. In particular, if the predictor is persistent and a forecasting relation does exist, then it must be the case that returns also have a highly persistent component. However, it is both theoretically unappealing as well as empirically unreasonable for realized returns to follow a nearly non-stationary process. In other words, if r_t has a persistent component then such a component must be very small, no matter how

[4]Suppose that in (1) $\phi < 1$ so that X_t is stationary. Then, using standard Central Limit Theorem arguments, we can show that $\sqrt{T}(\hat{\beta} - \beta) \to^d N\left(0, \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}\left(1 - \phi^2\right)\right)$, or $\hat{\beta} - \beta$ is approximately distributed as $N\left(0, \frac{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2}(1-\phi^2)}{T}\right)$. No matter how large this numerator is, as T increases, the variance of $\hat{\beta}$ will become arbitrarily close to zero.

large the sample size.

In this paper we assume that a predictive relation does prevail between returns and an explanatory variable.[5] In other words, we assume $\beta \neq 0$ in (1). We then investigate the properties of predictive regressions in light of the noisy nature of returns data as well as the persistence exhibited by many explanatory variables. For example, what are the implications for the properties of OLS estimates of β and corresponding t-statistics? Can we shed any additional light on the relation between the predictive regression's in-sample versus its out-of-sample forecasting properties? Answers to these questions cannot be obtained by assuming that the signal accumulates at the same rate as the sample size increases. Intuitively, what is required is an asymptotic theory in which the rate of information accumulation varies with the strength of the signal present in the data. We provide such a theory and show that it can help explain many of the stylized facts associated with predictive regressions.

All of these issues can be readily couched in the context of our predictive regression framework, (1)-(2). To emphasize the persistence of the forecasting variable X_t, we assume that $\phi = 1$ in (2) and is known.[6] To capture the noisiness of returns relative to the signal present in the data, recall that in the regression (1), the systematic component βX_t is the signal that helps us forecast the conditional mean of r_{t+1}, while the noise is the residual, ε_{t+1}. Heuristically, a small signal-noise ratio arises whenever σ_ε^2 is much larger than the regression's systematic component. Rescaling the regressor X_t does not remedy a small signal-noise ratio.[7] Also, since researchers have relied on explanatory variables expressed in a variety of units, for example, monthly *vs* quarterly observations, or logged *vs* non-logged variables, we will use a scale-invariant measure of the signal-noise ratio. In particular, we normalize both the signal and the noise by β.

[5] For example, lagged dividend yields can provide valuable information about future stock returns. This follows from Campbell and Shiller's log-linear framework where, appealing to the dynamic Gordon growth model, the log stock return, r_{t+1}, can be written as

$$r_{t+1} \approx k + \rho p_{t+1} + (1 - \rho) d_{t+1} - p_t$$

with p and d denoting the log stock price and log dividend, respectively, and k and ρ are parameters of the linearization.

[6] Alternatively, if ϕ is unknown but close to one, we may use the local to unity specification $\phi = 1 + \frac{c}{T}$ as in Cavanagh, Elliott and Stock (1995). In the Appendix, we show that allowing ϕ to be close to, but not exactly 1, will not affect our conclusions. The only difference will be that the results will depend on the nuisance parameter, c.

[7] In particular, suppose we rescale X_t by a constant K in an attempt to increase the signal. Or, equivalently, we consider

$$r_{t+1} = \mu + \beta^\dagger X_t^\dagger + \varepsilon_{t+1}$$

where $\beta^\dagger = \frac{\beta}{K}$ and $X_t^\dagger = X_t K$. Note that in order to preserve the posited relation, we must also rescale the regression coefficient, thereby keeping the variance of the systematic part unchanged.

It will also be convenient to rescale u_t as $u_t = \tau v_t$ where $\sigma_v^2 = \sigma_\varepsilon^2/\beta^2$ so that

$$\tau^2 = \frac{\sigma_u^2 \beta^2}{\sigma_\varepsilon^2}$$

represents the signal-noise ratio. If X_t is serially uncorrelated ($\phi = 0$) the signal-noise ratio is the familiar goodness-of-fit statistic R^2. However, since we model X_t with an autoregressive root of unity, τ^2 and R^2 are not identical but are related. In fact, we will see below that we can analyze the statistical properties of R^2 as a function of the signal-noise ratio.

In what follows we consider the predictive regression (1)-(2) augmented with:

$$\begin{pmatrix} \varepsilon_t \\ u_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \tau \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ v_t \end{pmatrix} \tag{3}$$

$$= \Upsilon w_t. \tag{4}$$

We assume that $w_t \equiv [\varepsilon_t, v_t]'$ is a martingale difference sequence with $E(w_t w_t' | w_{t-1}, ...) = [\sigma_\varepsilon^2 \;\; \sigma_{\varepsilon v}; \sigma_{\varepsilon v} \;\; \sigma_v^2] \equiv \Sigma$ and having finite fourth moments.

1.1 Preliminary Evidence

We are interested in the case where the signal-noise is small, or equivalently, the ratio $\tau = \sigma_u \beta / \sigma_\varepsilon$ is small. The summary statistics presented in Table 1 confirm that this case characterizes predictive regressions relying on a variety of commonly used explanatory variables.

The explanatory variables that we consider in addition to the log dividend yield (defined as the log annualized dividends minus the log price of the CRSP VW index) include the Treasury bill rate (defined as the log of one plus the three month Treasury bill rate), the long rate (defined as the log of one plus the yield of the ten year Treasury note), the default spread (defined as the log of one plus the BAA yield minus the log of one plus the AAA yield), the term spread (defined as the log of one plus the ten year Treasury note rate minus the log of one plus the three month Treasury bill rate), the real short rate (defined as log of one plus the three month Treasury bill rate minus the rate of inflation), as well as the relative short term rate of interest (defined as the log of one plus the one year Treasury bill rate minus its twelve month moving average). Annualized monthly data are used throughout.

Panel A of Table 1 provides the means and standard deviations of these variables over the entire sample period, 1927:1 - 1998:12, the sub-sample through approximately the end of the post-World

War II period, 1927:1 - 1949:12, the sub-sample subsequent to the post-World War II period, 1950:1 - 1998:12, and the first and second halves of this latter sub-sample, 1950:1 - 1979:12 and 1980:1 - 1998:12. For comparison purposes, we also provide these summary statistics for the equity premium (defined as the log of one plus the return of the CRSP VW index minus the log of one plus the one year Treasury bill rate). Notice how variable the equity premium is relative to all explanatory variables. Panel B presents the estimated slope coefficient $\hat{\beta}$ when the one-month ahead equity premium is regressed against these explanatory variables. We also provide estimates $\hat{\phi}$ of each explanatory variable's largest autoregressive root. In almost every case and across every sample period, the estimated ϕs are close to one, consistent with the persistent time series behavior of these variables. Finally, Panel C tabulates the corresponding R^2s which are seen to be close to zero and is indicative of the noisy nature of stock returns relative to the explanatory variables. This is confirmed by the estimated τs which are extremely small throughout.

1.2 Monte Carlo evidence

1.2.1 No correlation between ε_t and v_t

In light of the small signal-noise ratios that characterize predictive regressions using dividend yields and other explanatory variables, the following Monte Carlo experiments investigate the sensitivity of predictive regressions to varying signal-noise ratios. The system (1) - (4) is simulated for various T values, $T = 75,\ 200,\ 850$, and various τ values, $\tau = 1,\ 0.1,\ 0.01, ..., 1 \times 10^{-5}$. The case $T = 75$ is designed to correspond to a typical sample using annual data, $T = 200$ represents a typical sample using quarterly data, while $T = 850$ corresponds to a typical sample using monthly data. Without loss of generality, we will assume that $\beta = 1$ and set $\sigma_\varepsilon^2 = \sigma_v^2 = 1$, where ε_t and ν_t are *iid* normal variates with $\sigma_{\varepsilon v} = 0$. The system for each specification of (τ, T) is simulated $5,000$ times. At every simulation, we regress r_{t+1} on X_t, X_t on X_{t-1}, and r_t on r_{t-1}, resulting in estimates $\hat{\beta}$, $\hat{\phi}$, and $\hat{\phi}_2$, respectively. The estimated means and variances of these three OLS estimates are tabulated in Table 2.

Looking at Table 2A, the entries in the first row, $\tau = 1$ (i.e., $\sigma_u^2 = \sigma_\varepsilon^2 = 1$), show that the distribution of the β estimates is centered exactly at its true value of $\beta = 1$. However, as τ decreases, the β estimates worsen considerably, even for samples as large as $T = 850$. Since we have restricted ε_t and u_t to be independent, this result cannot reflect the small sample bias arising from the presence of a predetermined lagged variable, as discussed by Stambaugh (1999).

But if this sampling behavior of $\hat{\beta}$ does not reflect a small sample bias, how can we account for such poor performance even in reasonably large samples? Recall that traditional asymptotic theory predicts that the distribution of $\hat{\beta}$ will be centered around the true parameter value $\beta = 1$ but, as can be seen in Table 2A, this is clearly not the case as the signal-noise ratio diminishes. In fact, our simulations suggest that as τ decreases, the β estimates diverge further from the true value.

Table 2B documents the properties of the estimate of the autoregressive root in X_t. As expected, ϕ is estimated with downward bias, but this bias disappears as T increases. Notice that the estimate of ϕ is unaffected by the small signal-noise ratio, since its distribution is invariant to τ. Since X_t has a unit root and the relation (1) holds, then r_t must also have a unit root. Table 2C gives the estimates of the autoregressive root of r_t, ϕ_2. It is not surprising to find that for small values of τ the ϕ_2 estimates are not close to 1.[8]

As shown in Table 2D, inference is also problematic when τ is small. There we tabulate the mean of the distribution of the t-statistic for $H_o : \beta = 1$ versus $H_a : \beta = 0$. For τ small, the distributions under the null and under the alternative are both centered at zero, implying that the t-test will have almost no power even for samples of reasonable size (power equal to size). When the signal-noise ratio is small in a given sample, the R^2 must also, by definition, be small, but as T increases, R^2 must converge to one since we assume a relation prevailing between r_t and X_t. The mean of the R^2s obtained from the simulations are shown in Table 2E and we see that for the typical sample sizes encountered in practice the mean R^2s remain close to zero for small signal-noise ratios.

Lastly, in Table 2F we compare the out-of-sample forecast using the correct model, given in expression (1), to a forecast using the unconditional mean. For large values of τ, the forecast from the model outperforms the unconditional mean in a mean-squared error sense. As τ decreases, however, these forecasts produce similar results, even for reasonably large sample sizes.

1.2.2 Correlation between ε_t and v_t

If ε_t and v_t are assumed to be contemporaneously correlated, the OLS estimator of the slope coefficient, $\hat{\beta}$ will now be biased in small samples. As noted by Stambaugh (1999) and Cavanagh, Elliott, and Stock (1995), the size of this bias will depend on the magnitude of the fraction $\frac{cov(\varepsilon_t,u_t)}{var(u_t)}$

[8] To understand this result, notice that if τ is small, we can write $(1 - L)\, y_{t+1} = (1 + \theta L)\, r_{t+1}$ with θ very close to -1. In other words, y_{t+1} is very close to being white noise and it is known that under such circumstances unit root tests will have a size close to 1 (Perron (1988), Schwert (1989), and Pantula (1991)). Hence, it is not surprising that the null hypothesis of a unit root in r_t is rejected when σ_u^2 is small.

on the persistence of X_t. In our case, one can write $\frac{cov(\varepsilon_t,u_t)}{var(u_t)} = \frac{\tau\sigma_{\varepsilon v}}{\tau^2\sigma_v^2} = \frac{1}{\tau} \times \frac{\sigma_{\varepsilon v}}{\sigma_v^2}$ and it follows that a decrease in τ, all else being equal, leads to an increase in this small sample bias.

Our previous Monte Carlo experiments assumed that $\sigma_v = 1$ and we now set $\sigma_{\varepsilon v} = -0.62$ to capture the negative correlation prevailing between shocks to dividend yields and shocks to realized stock returns.[9] Table 3 provides the results when we repeat our Monte Carlo experiments with $\sigma_{\varepsilon v} = -0.62$. With few exceptions, these results are similar to the ones obtained assuming that $\sigma_{\varepsilon v} = 0$. From Table 3A we indeed see that the bias in $\hat{\beta}$ is exacerbated by the presence of contemporaneously correlated disturbances. Also, the means of the t-statistic under the null and alternative hypotheses are once again similar, and more importantly, do not diverge from each other with the increasing sample size, as would be required for a consistent statistical test. The t-statistic under the null now appears to be centered around 1.80. As we will show below, this shift occurs because of the correlation between the residuals and it is magnified by the small signal-noise ratio.

Taken together, the results of Tables 2 and 3 suggest that the estimation of and inference in predictive regressions may be adversely affected by the presence of a small signal relative to a large amount of noise. We now turn our attention to providing an asymptotic theory that verifies the sensitivity of predictive regressions to the prevailing signal-noise ratio.

2 Asymptotic Approximations when the Signal-Noise Ratio is Close to Zero

The Functional Central Limit Theorem can be used to provide better approximations of the small sample distributions of the various predictive regression statistics.

2.1 No correlation between ε_t and v_t

If $\sigma_{\varepsilon v} = 0$, then ε_t is uncorrelated with X_t and $\Sigma^{1/2}$ is diagonal. While this may not be a realistic assumption in the context of a predictive regression using dividend yields, we impose it to help explain the Monte Carlo simulation results of Table 2. The results we obtain will also serve as a basis for the more realistic case $\sigma_{\varepsilon v} \neq 0$.

Under this assumption, $(1/\sqrt{T}) \sum_{j=1}^{[sT]} w_j \Rightarrow \Upsilon\Sigma^{1/2}W(s)$ and $(1/\sqrt{T}) \sum_{j=1}^{[sT]} (w_j - \overline{w}) \Rightarrow \Upsilon\Sigma^{1/2}W^{\mu}(s)$, where $W(s) = [W_1(s)\ W_2(s)]'$ is a bivariate standard Weiner process on $D\,[0,1] \times D\,[0,1]$, $W^{\mu}(s) =$

[9] As estimated using data over the 1927:1 to 1998:12 sample period (See Table 1D).

$W(s) - \int_0^1 W(s)ds$, $t = [sT]$, and $\Rightarrow$ denotes convergence in distribution. To capture the small variance of u_t relative to ε_t and to explicitly link the behavior of the signal-noise ratio to the sample size, we write

$$\begin{aligned} \tau &\equiv \frac{\sigma_u \beta}{\sigma_\varepsilon} \\ &= \frac{1}{T^\alpha} \quad \text{where} \quad \alpha \geq 0. \end{aligned} \tag{5}$$

We can interpret the parameter α as a measure of the information in the conditioning variable βX_t relative to the noise ε_t:

$$\begin{aligned} \frac{\operatorname{Var}(\beta X_t)}{\operatorname{Var}(\varepsilon_t)} &= \frac{\sigma_u^2 t}{\sigma_\varepsilon^2}\beta^2 \\ &= \tau^2 t \\ &= \frac{t}{T^{2\alpha}} \\ &= sT^{1-2\alpha} \end{aligned}$$

for the increment $s = t/T$. The case $\alpha = 0$ corresponds the "usual" asymptotics where the signal-noise ratio τ is fixed. For $\alpha \in (0, 1/2)$, the information in the predictive variable will eventually overwhelm the noise and $\operatorname{Var}(X_t)/\operatorname{Var}(\varepsilon_t/\beta)$ diverges as $T \to \infty$. However, for $\alpha > 1/2$, the noise will engulf the signal and $\operatorname{Var}(X_t)/\operatorname{Var}(\varepsilon_t/\beta) \to 0$ as $T \to \infty$.

Given this parameterization, the following result obtains:

Proposition 1 *Under the assumptions above, if $\sigma_{\varepsilon v} = 0$ and $\tau = \frac{1}{T^\alpha}$, the OLS estimator $\hat{\beta}$ converges at rate $T^{-(1-\alpha)}$ to a functional of diffusion processes*:

$$T^{(1-\alpha)}\left(\hat{\beta} - \beta\right) \Rightarrow \frac{\int_0^1 W_2^\mu(s)dW_1(s)}{\int_0^1 (W_2^\mu(s))^2 ds} \quad \text{as} \quad T \to \infty. \tag{6}$$

More informally, we can write

$$\hat{\beta} \sim \beta + \tilde{Z}T^{\alpha-1}$$

where $\tilde{Z}$ is a mean zero normal random variable. If $\alpha < 1$, the variance of $\hat{\beta}$ will decrease as $T \to \infty$ and the estimator is consistent. Otherwise, $\hat{\beta}$ is inconsistent. Note, however, that $E(\hat{\beta}) = \beta$ since we assume $\sigma_{\varepsilon v} = 0$. The fact that we observed large simulation errors in Table 2A when computing $E(\hat{\beta})$ underscores the importance of Proposition 1.

While unusual, the above result is not surprising. Unusual, because $\hat{\beta}$ does not converge to β at rates $O(T^{1/2})$ or $O(T)$. In fact, the rate of convergence varies with α. However, this is not surprising, because we have parameterized the model so that α controls the rate at which the signal emanating from X_t accumulates. As α increases, the signal from X_t in a given sample decreases relative to the noise ε_t and the parameter β cannot be estimated precisely.

The following result helps us understand the behavior of t-statistics under varying signal-noise ratios:

Proposition 2 *Under the assumptions above, if* $\sigma_{\varepsilon v} = 0$ *and* $\tau = \frac{1}{T^\alpha}$ *then*

$$t = \frac{\hat{\beta} - \beta}{se(\hat{\beta})} \Rightarrow \frac{\int_0^1 W_2^\mu(s) dW_1(s)}{\left(\int_0^1 (W_2^\mu(s))^2 \, ds\right)^{1/2}} \quad \text{as} \quad T \to \infty.$$

The true value of β is unknown and it is customary to assume the null hypothesis of no predictability or $\beta = 0$. Under this null hypothesis, the test statistic is

$$\begin{aligned} t_{\hat{\beta}} &= \frac{\hat{\beta} - 0}{se(\hat{\beta})} \\ &= \frac{\hat{\beta} - \beta}{se(\hat{\beta})} + \frac{\beta - 0}{se(\hat{\beta})} \\ &= \frac{T^{1-\alpha}\left(\hat{\beta} - \beta\right)\left(\frac{1}{T^{1-\alpha}}\sum_{t=1}^T X_t^2\right)^{1/2}}{(\hat{\sigma}_\varepsilon^2)^{1/2}} + \frac{\left(T^{1-\alpha}\beta\right)\left(\frac{1}{T^{1-\alpha}}\sum_{t=1}^T X_t^2\right)^{1/2}}{(\hat{\sigma}_\varepsilon^2)^{1/2}}. \end{aligned}$$

By Proposition 2, the first term converges in distribution to a normally distributed random variable. However, the second term diverges for $\alpha < 1$, while for $\alpha \geq 1$ the second term converges to 0. As a result, if $\alpha < 1$, the statistic $t_{\hat{\beta}}$ provides a consistent test which can reject $\beta = 0$ when it is false. Unfortunately, for $\alpha \geq 1$ the distribution of $t_{\hat{\beta}}$ is the same for $\beta = 0$ as for the true value of β. In such a case, the test's power equals to its size and we are unable to reject $\beta = 0$ when it is false. While this perhaps extreme conclusion obtains for $\alpha \geq 1$, it also suggests that the t-test will have low power for values of α higher than 0.5.

2.2 Correlation between ε_t and v_t

If ε_t and v_t are assumed to be contemporaneously correlated then our preceding conclusions are only reinforced. The following Proposition generalizes our previous results to the case $\sigma_{\varepsilon v} \neq 0$:

Proposition 3 *Under the assumptions above, if $\sigma_{\varepsilon v} \neq 0$ and $\tau = \frac{1}{T^{\alpha}}$ then as $T \to \infty$*

$$T^{(1-\alpha)}\left(\hat{\beta}-\beta\right) \Rightarrow \left(1-\delta^2\right)^{1/2} \frac{\int_0^1 W_2^{\mu}(s)dW_{\perp}(s)}{\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds} + \delta \frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds}$$

$$t = \frac{\hat{\beta}-\beta}{se(\hat{\beta})} \Rightarrow \left(1-\delta^2\right)^{1/2} \frac{\int_0^1 W_2(s)dW_{\perp}(s)}{\left(\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} + \delta \frac{\int_0^1 W_2(s)dW_2(s)}{\left(\int_0^1 \left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}}$$

where $corr(\varepsilon_t, v_t) = \delta$ and $W_{\perp}(s)$ is a Wiener process obtained by projecting $W_1(s)$ on $W_2(s)$ with $E\left(W_{\perp}^2(s)\right) = 1-\delta^2$ and by construction, $W_{\perp}(s)$ and $W_2(s)$ are statistically independent.[10]

More informally, we can write

$$\hat{\beta} \sim \beta + \left(1-\delta^2\right)^{1/2} \tilde{Z}T^{(\alpha-1)} + \delta\tilde{R}T^{(\alpha-1)}$$

where $\tilde{Z}$ is a mean zero normal random variable while $\tilde{R}$ is a stochastic process with a defined density and a negative mean. Notice that if $\delta = 0$, the last term disappears in the above expressions and these results reduce to those obtained when $\sigma_{\varepsilon v} = 0$. All else being equal, the higher (lower) the correlation δ between the disturbances, the more (less) dominant is the last term. Since we know from Monte Carlo simulations that the density of the random variable $\tilde{R}$ has most of its mass on negative values, this generates a negative bias in finite samples, which is the result obtained by Stambaugh (1999) using different methods. The smaller the signal-noise ratio is, the larger this bias. For $\alpha > 1$, this bias does not disappear asymptotically and, as discussed above, the variance of $\hat{\beta}$ increases. Consistent with Cavanagh, Elliott and Stock (1995), to the extent that $\delta \neq 0$, it is not correct to use the standard normal distribution to assess the statistical significance of β. In fact, the smaller the signal-noise ratio is, the greater the deviation between the standard normal distribution and this appropriate distribution.

2.3 Monte Carlo–Once More

The results in Tables 2 and 3 were obtained by arbitrarily decreasing the value of τ. Tables 4 and 5 present the results of the same set of simulations but now for $\tau = \frac{1}{T^{\alpha}}$ and $\alpha = (0, 0.20, 0.5, 0.67, 1, 2)$. We assume that $\sigma_{\varepsilon,v} = 0$ in Table 4 while $\sigma_{\varepsilon,v} = -0.62$ in Table 5. As expected from the previous propositions, $\hat{\beta}$ is consistent for $\alpha < 1$ but is inconsistent for $\alpha > 1$ (see Tables 4A and

[10]This proposition can be generalized to accommodate error terms with more general autocorrelation and heteroskedasticity (Hansen (1992), Stock and Watson (1993)).

4B, and Tables 5A and 5B.). The t-statistic for the null $\beta = 1$ can also be seen to be inconsistent against the alternative $\beta = 0$ (Tables 4D and 5D).[11]

2.4 Spurious Regression?

A spurious regression arises whenever unrelated random walks are regressed against one another. Ferson, Sarkissian, and Simin (2000) contend that spurious regression is a concern in predictive regressions notwithstanding the fact that returns themselves are stationary. Their argument is based on the observation that returns can be decomposed into an unobserved expected return component and a noise component. To the extent that expected returns behave nearly like a random walk, a spurious regression arises when returns are regressed against a persistent predictor which is unrelated to the persistent expected return component.

In Appendix B we formally establish that such a spurious regression arises in the Ferson, Sarkissian, and Simin framework *only* if τ is constant and does not depend on T, that is, $\alpha = 0$. This is explicitly assumed by Ferson, Sarkissian, and Simin in their Monte Carlo simulations where the underlying coefficient of determination, R^2, is held constant throughout. Intuitively, in this case as T gets larger the persistent component of returns becomes increasingly dominant. An implication of their assumption, however, is that returns themselves will increasingly behave like a random walk which is neither theoretically appealing nor empirically tenable. In other words, to the extent that realized returns are stationary then there is little concern that predictive regressions are spurious.

3 The Explanatory Power of Predictive Regressions

Recent work by Bossaerts and Hillion (1999), using statistical model selection criteria applied to a variety of explanatory variables, and Goyal and Welch (1999), relying exclusively on dividend yields, demonstrate that predictive regressions have some in-sample predictability but exhibit little or no out-of-sample predictability. While Bossaerts and Hillion suggest that these results may be due to model nonstationarity, in this section we argue that the excessive noisy nature of returns relative to the explanatory variables can explain both the apparent presence of in-sample predictability and the failure to detect out-of-sample forecasting power.

[11]Recall from Proposition 2 that the inconsistency of the t-statistic does not depend on the choice of the alternative hypothesis.

3.1 In-Sample

When estimating and testing one-period ahead predictive regressions, the t-statistic is usually only marginally significant and the coefficient of determination R^2 is low. However, given the posited relation assumed to prevail between r_{t+1} and X_t, one would expect when relying on normal asymptotics the resultant R^2 to increase to one with increasing sample size. In fact, in the one-period ahead predictive regression we can establish the following result:

Proposition 4 *Under the assumptions above, if $\tau = \frac{1}{T^\alpha}$, then as $T \to \infty$*

$$R^2 \to^p \begin{cases} 1 & \alpha < 1/2 \\ 0 & \alpha > 1/2 \end{cases}$$

where $\to^p$ denotes convergence in probability. For the borderline case $\alpha = 1/2$, $R^2 = O_p(1)$.

For a sufficiently small signal-noise ratio, an increase in the number of observations will not result in an increase in R^2, even if a relation prevails between the two variables. Enlarging the number of observations (going, for example, from a yearly to a monthly frequency) may not result in a higher R^2. In fact, we typically see that the R^2s in the monthly predictive regressions are considerably lower than the R^2s obtained in predictive regressions using yearly data.

To illustrate the above proposition, we turn to the simulation results in Tables 2E, 3E, 4E, and 5E. The first two tables present the simulation results for R^2 assuming τ fixed with $\sigma_{\varepsilon,v} = 0$ and $\sigma_{\varepsilon,v} = -0.62$, respectively. No matter what the correlation is, for a large τ we see that R^2 increases to one as the sample size increases. However, for very small values of τ, R^2 appears to converge to zero. This result cannot be explained by usual fixed τ asymptotics but is in accord with the above Proposition. Tables 4E and 5E presents the results for various α values and $\sigma_{\varepsilon,v} = 0$ and $\sigma_{\varepsilon,v} = -0.62$, respectively. We can clearly see that for $a < 1/2$, R^2 increases with the sample size but for $\alpha > 1/2$, R^2 converges to zero.

Bossaerts and Hillion argue that the lack of in-sample predictability of predictive regressions cannot be due to a small signal-noise ratio. In particular, they show that if one assumes that the $R^2 = 0.06$ observed in their empirical regressions is the "true" coefficient of determination, then the power of a corresponding t-test is fairly high. A reported R^2, however, is but an estimate with a finite-sample distribution. Moreover, as shown above, for small values of α the coefficient of determination does not converge to one in probability. Figure 1 graphically displays the distribution of R^2 for various τ and α values. The upper panels provide the entire distribution of R^2, while the lower panels show only the $0.5, 5, 50, 95$, and 99.5 percentiles, respectively. For a median value of $R^2 = 0.06$ (denoted by the dashed lines), we can trace 90% and 99% confidence intervals for R^2 corresponding to (0,0.17) and (0,0.20), respectively (denoted by dashed-dotted lines). As a result, referring to Bossaerts and Hillion's Figure 1, we conclude that a lack of in-sample predictability may very well be due to the t-statistic's lack of power. Of course, this is nothing but a restatement of our conclusions in Proposition 2.

3.2 Out-of-Sample

Given that the slope coefficient β is not estimated precisely, how good are the out-of-sample forecasts of one-period ahead predictive regressions? In the finance literature, the results of out-of-sample forecasting exercises are often seen as the most relevant measure of the success of a particular model. In light of our previous discussion, we do not expect to be able to forecast returns with great accuracy, notwithstanding the relation assumed to prevail between returns and a forecasting variable like the dividend yield. Consistent with this, Goyal and Welch (1999) point out that equity forecasts produced from a predictive regression model using annual data do not perform any better than the unconditional mean return.

We now prove that if the signal-noise ratio is small, forecasting using the estimated predictive regression model will not do better than the simple unconditional mean. First, we show that traditional asymptotics with fixed τ cannot give us insights into this problem. Then, we turn attention to our alternative asymptotic theory and derive analytic results that help explain the results of Goyal and Welch as well as the results of our simulations. In what follows, we compare two competing long-run forecasts: $\overline{r} = \frac{1}{T}\sum_{t=1}^{T} r_t$ and $\hat{r}_{T+k|T} = \hat{\mu} + \hat{\beta}x_T$, where $k = [\kappa T]$, and $\kappa \in (0,1)$.

For a fixed τ, we can show that both forecasts are asymptotically unbiased, or

$$E\left(T^{-1/2}\left(r_{T+k} - \overline{r}\right)\right) \to 0$$
$$E\left(T^{-1/2}\left(r_{T+k} - \hat{r}_{T+k|T}\right)\right) \to 0.$$

More importantly, the asymptotic variances are:

$$E\left(T^{-1}\left(r_{T+k} - \overline{r}\right)^2\right) \to \tau^2\beta^2(\kappa + 1/3)$$
$$E\left(T^{-1}\left(r_{T+k} - \hat{r}_{T+k|T}\right)^2\right) \to \tau^2\beta^2\kappa.$$

Therefore, we can conclude that asymptotically,

$$MSE(\overline{r}) > MSE(\hat{r}_{T+k|T}).$$

While this result is to be expected, it is inconsistent with our simulation results which suggest that for small τ the MSEs from both forecasts are almost identical, even for relatively large sample sizes. This fact, however, can be captured using our alternative asymptotic methods. As the following Proposition demonstrates, under this parameterization, $\hat{r}_{T+k|T}$ does not always produce superior forecasts to $\overline{r}$.

Proposition 5 *Under the assumptions above, suppose* $\tau = \frac{1}{T^{\alpha}}$ *and* $k=[\kappa T]$ *where* κ *is a fixed number. Let* $\overline{r}$ *be the sample mean of* r_t *and let* $\hat{r}_{T+k|T}=\hat{\alpha}+\hat{\beta}X_T$*. Then, as* $T\to\infty$*, both forecasts are asymptotically unbiased for all values of* α*:*

$$\begin{array}{ll} E\left[T^{-(1/2-\alpha)}\left(r_{T+k}-\overline{r}\right)\right]\to 0 \text{ and } E\left[T^{-(1/2-\alpha)}\left[r_{T+k}-\hat{r}_{T+k|T}\right]\right]\to 0 & ,\ \alpha<1/2 \\ E\left(\left(r_{T+k}-\overline{r}\right)\right)\to 0 \text{ and } E\left[\left[r_{T+k}-\hat{r}_{T+k|T}\right]\right]\to 0 & ,\ \alpha\geq 1/2. \end{array}$$

However,

$$\begin{array}{ll} E\left(T^{-(1-2\alpha)}\left(r_{T+k}-\overline{r}\right)^2\right)\to\beta^2(\kappa+1/3) & ,\ \alpha<1/2 \\ E\left(\left(r_{T+k}-\overline{r}\right)^2\right)\to\beta^2(\kappa+1/3)+1 & ,\ \alpha=1/2 \\ E\left(\left(r_{T+k}-\overline{r}\right)^2\right)\to 1 & ,\ \alpha>1/2. \end{array}$$

and

$$\begin{array}{ll} E\left(T^{-(1-2\alpha)}\left[r_{T+k}-\hat{r}_{T+k|T}\right]^2\right)\to\beta^2\kappa & ,\ \alpha<1/2 \\ E\left(\left[r_{T+k}-\hat{r}_{T+k|T}\right]^2\right)\to\beta^2\kappa+1 & ,\ \alpha=1/2 \\ E\left(\left[r_{T+k}-\hat{r}_{T+k|T}\right]^2\right)\to 1 & ,\ \alpha>1/2 \end{array}$$

Therefore,

$$\begin{array}{ll} MSE(\overline{r}) > MSE(\hat{r}_{T+k|T}) & ,\ \alpha<1/2 \\ MSE(\overline{r}) = MSE(\hat{r}_{T+k|T}) & ,\ \alpha\geq 1/2. \end{array}$$

Notice that in the case $\alpha > 1/2$, we have $MSE(\overline{r}) = MSE(\hat{r}_{T+k|T})$. In other words, if the signal-noise ratio is extremely low then forecasts from the true model will not necessarily outperform the unconditional mean return.

In Table 6, we compare the forecasting power of our sampled explanatory variables. Panel A displays the percentage *increase* in MSE from using a given predictor instead of the unconditional mean return. We confirm the findings of Goyal and Welch (1999) that the log dividend yield not only fails to significantly outperform the simple mean return, but in some periods it actually produces inferior forecasts. Strikingly, the remaining explanatory variables do not forecast much better. In particular, no single forecaster uniformly dominates the others across all sub-samples. The log dividend yield performs particularly poorly in the last sub-sample period. The default spread yields a high MSE during the 1927-1949 and 1980-1998 sample periods. The term spread performs particularly poorly during the 1950-1979 sample period while the real interest rate's forecasting performance deteriorates during the 1927-1949 sample period. In general, the best forecaster is the relative rate. In Panel B, we compare the predictive ability of our forecasters by re-estimating the model at each sampled time point. Such a "rolling-estimation," although not a formal test, allows us to check whether the lack of forecastability is due to an unstable relation or to a lack of signal.[12] Since the results in Panels A and B are in general agreement, we conclude that even if instability is present it is not the only source of the lack of predictability.

4 Estimation of α

In the previous sections we have argued that the signal-noise ratio must be taken into account when conducting estimation, inference, and forecasting with predictive regressions. Under the proposed asymptotic approximations, the rates of convergence of all required statistics depend on the rate of signal accumulation. This feature is not only intuitively appealing but also desirable. It provides us with guidelines, or boundaries, delineating how much forecasting power to expect from a particular posited explanatory variable. In this section we provide a quick and simple method of estimating α.

Notice that for $t = T$ we have $\frac{\sigma_u^2}{\sigma_\varepsilon^2}\beta^2 = T^{-2\alpha}$. Therefore, taking logs and re-arranging the above expression, we obtain

$$\alpha = \frac{\log \sigma_e - \log \sigma_u - \log \beta}{\log T}. \tag{7}$$

If we replace the population moments with their sample counterparts and use the OLS estimate of β, we obtain a consistent estimate of α, say $\hat{\alpha}^{OLS}$. Unfortunately, given the unusual rates of convergence, an analytic form of the limiting distribution of $\hat{\alpha}^{OLS}$ is difficult to derive. Moreover, given the slow rates of convergence of $\hat{\beta}$ discussed previously, an estimate of α constructed using (7)

[12] Note that some formal tests for instability are based on such rolling-regression estimation schemes (Viceira (1997)). Therefore, this simple diagnostic should allow us to detect time-varying relation, if it were present.

will have a small sample bias. To remedy the shortcomings of $\hat{\alpha}^{OLS}$, we construct an estimator of α that will be median-unbiased and around which we can define valid confidence intervals. Median-unbiased estimators are well-known in statistics (see Lehmann (1959)) and have recently been used in econometrics by Andrews (1993) and Stock (1991).

We propose to estimate α using a method that produces a median-unbiased estimate, $\hat{\alpha}^{MU}$, and corresponding confidence intervals with exact coverage even in small samples. The procedure involves correcting the bias in $\hat{\alpha}^{OLS}$ by inverting its distribution. More formally, we follow Andrews (1993). Suppose we have the model in (1)-(5) and the estimator $\hat{\alpha}^{OLS}$ defined by (7). Notice that the distribution of $\hat{\alpha}^{OLS}$ depends only on α and δ, the correlation between ε_t and u_t. Since δ can be estimated consistently, we will assume that it is known. The median of the distribution of $\hat{\alpha}^{OLS}$, denoted by $M(\alpha)$, will be only a function of α and is defined by $P\left(\hat{\alpha}^{OLS} \leq M(\alpha)\right) = 0.5$. Suppose that $M(\alpha)$ is strictly increasing on the parameter space Π, say $(0, 1.2]$. Then, $\hat{\alpha}^{MU}$ is a median-estimator of α if

$$\hat{\alpha}^{MU} = \begin{cases} M^{-1}\left(\hat{\alpha}^{OLS}\right) & \hat{\alpha}^{OLS} \leq M\left(\max_{\alpha} \Pi\right) \\ \max_{\alpha} \Pi & \hat{\alpha}^{OLS} > M\left(\max_{\alpha} \Pi\right). \end{cases} \tag{8}$$

The truncation $\max_{\alpha} \Pi$ is necessary since the estimated value may fall outside the predetermined parameter space Π. [13]

The construction of $100 \times (1-p)\,\%$ confidence intervals follows in the same fashion, given the definition of the quantiles $Q_{p1}(\alpha)$ and $Q_{p2}(\alpha)$, $P(Q_{p1}(\alpha) \leq \hat{\alpha}^{OLS} \leq Q_{p2}(\alpha)) = 1 - p$, where $p1 + p2 = p$. For example, if we consider the 99% centered confidence interval, we have $P(Q_{0.005}(\alpha) \leq \hat{\alpha}^{OLS} \leq Q_{0.995}(\alpha)) = 0.99$. Replacing $\hat{\alpha}^{MU}$, $M(\alpha)$, and $M^{-1}\left(\hat{\alpha}^{OLS}\right)$ in definition (8) by α_{LB}, $Q_{0.995}(\alpha)$, $Q_{0.995}^{-1}\left(\hat{\alpha}^{OLS}\right)$, we obtain a lower bound on the 99% confidence interval of $\hat{\alpha}^{MU}$. Similarly, replacing $\hat{\alpha}^{MU}$, $M(\alpha)$, and $M^{-1}\left(\hat{\alpha}^{OLS}\right)$ in definition (8) by α_{UB}, $Q_{0.005}(\alpha)$, $Q_{0.005}^{-1}\left(\hat{\alpha}^{OLS}\right)$, we obtain an upper bound on the 99% confidence interval of $\hat{\alpha}^{MU}$.

Inverting the distribution of a test statistic in order to obtain $\hat{\alpha}^{MU}$, α_{LB}, and α_{UB} is easily implemented. Since we know the data generating process in (1)-(5), we can simulate the distributions of $\hat{\alpha}^{OLS}$ as a function of the true parameter α. Therefore, we can find the quantiles $Q_{0.005}(\alpha)$, $M(\alpha)$, and $Q_{0.995}(\alpha)$ and invert them. As an illustration of the method, consider the predictive regression using the log dividend yield over the sample period 1927:1-1998:12. Using the OLS regression results and (7) we obtain $\hat{\alpha}^{OLS} = 0.72$ (Table 7B). Given the estimate $\hat{\delta} = -0.62$, we

[13] In practice, this is not a limitation, since the parameter space can be set to be arbitrarily large. In the application of this procedure, we set $\Pi = (0, 1.2]$ and the truncation is never used to find $\hat{\alpha}^{MU}$.

simulate the distribution of $\hat{\alpha}^{OLS}$ as a function of α for $T = 850$. The 0.5th, 50th, and 99.5th quantiles of the distribution are plotted in Figure 2 as a function of α. A graphical inversion of the quantiles, shown by dashed lines, produces $\left(\alpha_{LB},\ \hat{\alpha}^{MU},\ \alpha_{UB}\right) = (0.48,\ 0.59,\ 0.73)$.

The advantages of the proposed estimation procedure are clear. First, we obtain an unbiased estimator of α. Second, the confidence intervals have the correct coverage by definition even in small samples and can be used for accurate hypothesis testing. Third, the method is easy to implement. Fourth, the impartiality of the median unbiased estimator is particularly attractive. We do not need to appeal to asymptotic approximations (in the Frequentist approach) nor to various priors (in the Bayesian approach) in order to produce the statistic of interest. The only drawback of the procedure is that it necessitates the simulation of the data generating process in (1-5), but that is a small cost given the simplicity of the model.[14]

We use the estimator $\hat{\alpha}^{MU}$ to measure the signal strength of our sampled explanatory variables over various sample periods. These estimates are displayed in Table 7A along with the corresponding correctly-sized 99% confidence intervals around $\hat{\alpha}^{MU}$. For comparison purposes, the biased least squares estimates $\hat{\alpha}^{OLS}$ are provided in Table 7B.

A number of interesting results emerge from Table 7A.[15] First, as expected, the estimates of α are fairly high for all the regressors and range between 0.26 to 0.95. The α confidence intervals of all the explanatory variables in all sample periods fail to include $\alpha = 0$ but most do include α values of 0.5 or higher consistent with their signal-noise ratios being low. Given these results, it is not surprising that none of these explanatory variables can adequately explain the fluctuations in returns in- or out-of-sample. Second, there is no explanatory variable that performs worse than the others across all sub-samples. However, the dividend yield, the long rate and the default spread have consistently higher α estimates than the other predictors. The dividend yield lacks signal strength during the 1927-1950 and particularly the 1980-1998 sample periods, but performs quite well otherwise. Interestingly, the relative rate's estimate of α is consistently lower than that of other variables across all sample periods. Moreover, it is significantly lower than 0.5 throughout, thus explaining the observed superiority of this forecaster in Table 6. Third, we note that there is a considerable variation in the signal strength from one sample period to another with the largest

[14] There is another, more technical drawback of the median unbiased estimator. There might be better median-unbiased estimators than the one proposed here, i.e. estimators that are more efficient in exploiting the information in the data. We could not find any known results about the optimality of median unbiased estimators.

[15] To emphasize the usefulness of α as a measure of signal strength, we note that the conclusions that we drew from Tables 1 and 6 correspond closely to those from Table 7.

variation across periods coming from the dividend yield. Again, the only predictor that has a fairly constant α estimate is the relative rate. In sum, the relative rate is the only predictor that can outperform the unconditional mean of returns. The other predictors simply do not have enough signal.[16] Interestingly, Campbell (1991) reaches a similar conclusion using more heuristic arguments.

5 Conclusions

Stock returns are extremely noisy when compared to forecasting variables like the dividend yield commonly used in predictive regressions. As a result, standard statistical tests and procedures will have difficulty in detecting and properly gauging predictability when it is actually present in the data.

Researchers typically use persistent conditioning variables in forecasting stock returns. To better understand the results of predictive regressions in finite samples, this paper provides an alternative asymptotic approximation, where the rate at which information accumulates with the increasing sample size is explicitly controlled. The proposed framework allows us to analytically explore the statistical boundaries of predictive regressions in finite samples, when the signal-noise ratio is small. Although our analysis assumes that the predictors are persistent variables, this need not be the case. In fact, similar results can easily be derived for stationary variables if we parameterize the signal-noise ratio to be decreasing with the sample size.

The novelty of our parameterization is that the rate of information accumulation explicitly influences the rate of convergence and the behavior of various statistical properties of predictive regressions. Using these results, a researcher can gauge the informativeness of a particular forecasting variable relative to its noise and determine whether reliable estimation, inference, and forecasting can be expected in predicting stock returns. In a bivariate framework, we use our methodology to demonstrate that even if a forecasting relation does exist between returns and some commonly used predictors, the relative rate is the only predictor to deliver some forecasting gains when compared to the unconditional mean of returns.

[16] Using sample moments in (7) is not the only method to estimate α. An alternative method would be to invert the distribution of a sample statistic that depends on α. For example, one might want to invert the distributions of the R^2 statistic, shown in Figure 1. This method, proposed by Andrews (1993) and Stock (1991) will produce a median unbiased estimate of the parameter as well as valid confidence intervals.

Tables and Figures

Table 1: Summary Statistics

A: Means and Standard Deviations

	1927:1-1998:12		1927:1-1949:12		1950:1-1998:12		1950:1-1979:12		1980:1-1998:12	
	mean	stdev	mean	stdev	mean	stdev	mean	stdev	mean	stdev
EP	6.46	19.20	4.96	26.52	7.16	14.58	6.31	13.89	8.51	15.63
DY	-3.07	0.33	-2.85	0.27	-3.18	0.30	-3.11	0.25	-3.28	0.35
TBL	3.71	0.91	0.90	0.38	5.03	0.83	3.93	0.65	6.78	0.84
LONG	5.35	0.87	2.71	0.15	6.59	0.83	5.06	0.59	9.01	0.67
DSPR	1.14	0.21	1.62	0.29	0.92	0.12	0.78	0.09	1.13	0.14
TSPR	1.64	0.37	1.81	0.31	1.56	0.40	1.13	0.29	2.23	0.47
RTBL	0.60	1.94	-0.44	3.04	1.08	1.08	-0.02	1.05	2.81	0.93
RR	0.01	0.29	-0.05	0.17	0.04	0.33	0.17	0.25	-0.16	0.42

B: Estimates of β and highest AR root, ϕ

	1927:1-1998:12		1927:1-1949:12		1950:1-1998:12		1950:1-1979:12		1980:1-1998:12	
	β	ϕ	β	ϕ	β	ϕ	β	ϕ	β	ϕ
DY	0.01	0.98	0.01	0.94	0.01	0.99	0.03*	0.98	-0.00	0.99
	(0.90)	(-1.35)	(0.45)	(-2.30)	(1.43)	(-0.47)	(3.04)	(-1.96)	(-0.44)	(0.70)
TBL	-0.95	0.98	-2.83	0.96	-1.73*	0.98	-3.27*	0.98	-2.01	0.97
	(-1.32)	(-1.96)	(-0.67)	(-2.14)	(-2.40)	(-2.29)	(-2.88)	(-0.56)	(-1.63)	(-1.33)
LONG	-0.72	1.00	-11.57	0.99	-1.30	1.00	-3.12*	1.00	-2.52	0.99
	(-0.96)	(-1.26)	(-1.06)	(-1.39)	(-1.80)	(-1.65)	(-2.51)	(0.77)	(-1.62)	(-0.77)
DSPR	2.29	0.98	2.54	0.98	6.36	0.97	10.81	0.97	3.42	0.97
	(0.75)	(-2.57)	(0.45)	(-1.44)	(1.31)	(-3.15)	(1.32)	(-2.81)	(0.47)	(-2.27)
TSPR	1.73	0.90*	1.67	0.95	1.86	0.89*	3.24	0.88*	1.38	0.87
	(0.98)	(-4.61)	(0.32)	(-2.08)	(1.23)	(-4.06)	(1.26)	(-3.40)	(0.62)	(-2.66)
RTBL	0.26	0.71*	0.09	0.71*	0.81	0.71*	0.48	0.54*	1.59	0.67*
	(0.77)	(-6.86)	(0.17)	(-3.85)	(1.46)	(-5.68)	(0.69)	(-5.64)	(1.43)	(-4.68)
RR	-3.24	0.75*	8.36	0.63*	-4.64*	0.76*	-5.01	0.78*	-4.62	0.77*
	(-1.44)	(-8.07)	(0.86)	(-5.02)	(-2.57)	(-6.56)	(-1.72)	(-5.32)	(-1.90)	(-3.85)

Table 1 (Cont'd): Summary Statistics

C: $R^2 * 100$ and τ

	1927:1-1998:12		1927:1-1949:12		1950:1-1998:12		1950:1-1979:12		1980:1-1998:12	
	R^2	τ	R^2	τ	R^2	τ	R^2	τ	R^2	τ
DY	0.09	7.82e-003	0.07	9.58e-003	0.35	1.50e-002	2.52	5.49e-002	0.09	5.11e-003
TBL	0.20	1.03e-002	0.16	1.66e-002	0.97	2.72e-002	2.26	4.22e-002	1.17	3.75e-002
LONG	0.11	2.60e-003	0.41	1.15e-002	0.55	7.30e-003	1.73	1.07e-002	1.16	1.88e-002
DSPR	0.06	5.48e-003	0.08	6.77e-003	0.29	1.21e-002	0.49	1.53e-002	0.10	8.09e-003
TSPR	0.11	1.90e-002	0.04	9.89e-003	0.26	2.97e-002	0.44	4.06e-002	0.17	2.61e-002
RTBL	0.07	2.31e-002	0.01	8.93e-003	0.36	5.41e-002	0.13	3.49e-002	0.90	8.07e-002
RR	0.24	3.43e-002	0.27	4.59e-002	1.12	7.17e-002	0.82	6.20e-002	1.58	8.55e-002

D: Correlation between ϵ_t and u_t

	1927:1-1998:12	1927:1-1949:12	1950:1-1998:12	1950:1-1979:12	1980:1-1998:12
	corr(ϵ_t,u_t)	corr(ϵ_t,u_t)	corr(ϵ_t,u_t)	corr(ϵ_t,u_t)	corr(ϵ_t,u_t)
DY	-0.62	-0.78	-0.47	-0.40	-0.62
TBL	-0.07	0.05	-0.17	-0.19	-0.16
LONG	-0.18	-0.12	-0.26	-0.21	-0.34
DSPR	-0.27	-0.42	0.04	0.08	0.00
TSPR	0.02	-0.07	0.08	0.14	0.03
RTBL	-0.04	-0.08	0.04	0.04	0.06
RR	-0.08	0.09	-0.18	-0.18	-0.17

Notes: All series with the exception of DY are expressed in annualized percentage points. DY is the log dividend yield. For consistency with the previous literature, we use log returns throughout this paper (c.f. Campbell et. al (1997)). The results are unchanged when simple returns are used. The following acronyms will be used in the tables below: EP is the log of one plus the return on the value weighted CRSP portfolio minus log of 1 plus the yield of the three month Treasury bill, DY is the log annualized dividends minus the log price of the value weighted CRSP index, TBL is the log of one plus the yield of the three month Treasury bill, LONG is the log of one plus the yield of the ten year Treasury note, DSPR is the log of one plus the BAA yield minus the log of one plus the AAA yield, TSPR is the log of one plus the ten year Treasury note rate minus the log of one plus the three month Treasury bill rate, RTBL is the log of one plus the three month Treasury bill rate minus the log of one plus the inflation rate, and RR is the log of one plus the one year Treasury bill rate minus its twelve month moving average. In Tableau B, the β coefficients are tested for being equal to zero using a t-statistic, whereas the ϕ coefficients are tested for being equal to 1, using an Augmented Dickey-Fuller (ADF) test. The symbol "*" denotes significance at the 5 percent level of the respective test. The parameter τ, defined in the text as $\tau = \sigma_u * \beta / \sigma_\epsilon$ is estimated using the sample analogues of the standard deviations and the least squares estimate of β.

Table 2A

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 850	
τ	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
1	1.00	1.99e-003	1.00	2.69e-004	1.00	1.51e-005
1.0e-001	1.00	1.90e-001	1.00	2.74e-002	1.00	1.50e-003
1.0e-002	0.97	1.93e+001	0.98	2.80e+000	1.00	1.48e-001
1.0e-003	0.96	2.03e+003	0.70	2.72e+002	0.99	1.45e+001
1.0e-004	0.09	1.98e+005	1.60	2.89e+004	0.87	1.50e+003
1.0e-005	-28.80	1.97e+007	25.16	2.82e+006	2.97	1.49e+005

Table 2B

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 850	
τ	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
1	0.98	1.67e-003	0.99	2.56e-004	1.00	1.34e-005
1.0e-001	0.98	1.74e-003	0.99	2.32e-004	1.00	1.46e-005
1.0e-002	0.98	1.66e-003	0.99	2.51e-004	1.00	1.49e-005
1.0e-003	0.98	1.73e-003	0.99	2.74e-004	1.00	1.38e-005
1.0e-004	0.98	1.59e-003	0.99	2.58e-004	1.00	1.37e-005
1.0e-005	0.98	1.62e-003	0.99	2.32e-004	1.00	1.34e-005

Table 2C

The OLS estimator of the root in r_t

	T= 75		T= 200		T= 850	
τ	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
1	0.82	1.45e-002	0.92	3.01e-003	0.98	2.05e-004
1.0e-001	0.09	2.00e-002	0.21	2.17e-002	0.50	3.05e-002
1.0e-002	-0.01	1.35e-002	-0.00	4.89e-003	0.01	1.30e-003
1.0e-003	-0.01	1.34e-002	-0.00	4.96e-003	-0.00	1.15e-003
1.0e-004	-0.01	1.34e-002	-0.00	5.11e-003	-0.00	1.16e-003
1.0e-005	-0.01	1.31e-002	-0.00	5.02e-003	-0.00	1.17e-003

Notes: See next page.

Table 2D

Mean of t-stat under Null and Alternative

τ	T= 75 null	T= 75 alt	T= 200 null	T= 200 alt	T= 850 null	T= 850 alt
1	0.04	28.03	-0.00	75.10	0.00	321.05
1.0e-001	0.00	2.85	0.00	7.53	-0.03	32.03
1.0e-002	-0.02	0.28	-0.03	0.73	0.00	3.21
1.0e-003	0.01	0.03	-0.03	0.06	-0.00	0.32
1.0e-004	-0.00	-0.00	0.02	0.01	0.01	0.04
1.0e-005	-0.00	-0.01	0.03	0.01	0.03	0.01

Table 2E

Mean of R^2

τ	T= 75	T= 200	T= 850
1	0.88	0.95	0.99
1.0e-001	0.12	0.22	0.51
1.0e-002	0.02	0.01	0.01
1.0e-003	0.01	0.01	0.00
1.0e-004	0.01	0.01	0.00
1.0e-005	0.01	0.01	0.00

Table 2F

Comparison of MSE's from $\bar{r}$ and $\hat{r}_{T+k|T}$

	T= 75			T= 200			T= 850		
τ	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio
1	0.375	0.054	6.968	0.380	0.049	7.719	0.376	0.050	7.467
1.0e-001	0.017	0.014	1.208	0.009	0.005	1.625	0.005	0.002	2.976
1.0e-002	0.014	0.014	0.986	0.005	0.005	0.996	0.001	0.001	1.028
1.0e-003	0.014	0.014	0.968	0.005	0.005	0.995	0.001	0.001	0.999
1.0e-004	0.013	0.014	0.980	0.005	0.005	0.989	0.001	0.001	0.995
1.0e-005	0.013	0.013	0.976	0.005	0.005	0.989	0.001	0.001	0.998

Notes: The simulated system is: $r_{t+1} = \mu + \beta X_t + \varepsilon_{t+1}$, $X_{t+1} = \mu_x + \phi X_t + u_{t+1}$ var(ε_t)=1, var(u_t)=τ^2 $\times$ var(ε_t). Since we are considering very persistent predictors, we let $\phi = 1$ but similar results obtain for ϕ close to unity. Without loss of generality, we let $\beta = 1$. Note that, given the normalization used in the text, the simulations are invariant with respect to β. τ is a fixed small number. The correlation between u_t and ε_t is zero, so there is no bias. The system is simulated 5000 times, for each specification of (τ, T). The first and second moments in Tables A-C are calculated by using sample analogues. Table F compares the out-of-sample forecast of the unconditional vs. the conditional mean of r_t, k periods ahead, where k=0.01 $\times$ T. Very similar results are obtained for fractions other than 0.01.

Table 3A

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 850	
τ	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
1	1.04	2.42e-003	1.02	3.64e-004	1.00	1.97e-005
1.0e-001	1.44	2.44e-001	1.17	3.56e-002	1.04	1.98e-003
1.0e-002	5.31	2.44e+001	2.65	3.59e+000	1.39	1.91e-001
1.0e-003	43.80	2.37e+003	17.53	3.46e+002	4.99	2.05e+001
1.0e-004	426.62	2.43e+005	160.49	3.30e+004	39.30	1.89e+003
1.0e-005	4347.70	2.52e+007	1651.65	3.45e+006	407.97	2.05e+005

Table 3B

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 850	
τ	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
1	0.98	1.64e-003	0.99	2.54e-004	1.00	1.38e-005
1.0e-001	0.98	1.68e-003	0.99	2.54e-004	1.00	1.32e-005
1.0e-002	0.98	1.69e-003	0.99	2.43e-004	1.00	1.44e-005
1.0e-003	0.98	1.58e-003	0.99	2.31e-004	1.00	1.41e-005
1.0e-004	0.98	1.57e-003	0.99	2.26e-004	1.00	1.42e-005
1.0e-005	0.98	1.70e-003	0.99	2.42e-004	1.00	1.47e-005

Table 3C

The OLS estimator of the root in r_t

	T= 75		T= 200		T= 850	
τ	$E(\hat{\phi_2})$	$Var(\hat{\phi_2})$	$E(\hat{\phi_2})$	$Var(\hat{\phi_2})$	$E(\hat{\phi_2})$	$Var(\hat{\phi_2})$
1	0.77	1.98e-002	0.90	5.14e-003	0.97	3.91e-004
1.0e-001	0.08	2.00e-002	0.20	2.09e-002	0.49	3.10e-002
1.0e-002	-0.02	1.31e-002	-0.00	5.11e-003	0.01	1.35e-003
1.0e-003	-0.01	1.32e-002	-0.01	5.04e-003	-0.00	1.22e-003
1.0e-004	-0.01	1.32e-002	-0.00	4.85e-003	-0.00	1.22e-003
1.0e-005	-0.01	1.32e-002	-0.01	4.86e-003	-0.00	1.18e-003

Notes: See next page.

Table 3D

Mean of t-stat under Null and Alternative

	T= 75		T= 200		T= 850	
τ	null	alt	null	alt	null	alt
1	1.87	29.35	1.83	76.24	1.78	320.88
1.0e-001	1.83	3.75	1.79	8.48	1.77	33.06
1.0e-002	1.80	1.22	1.80	1.71	1.79	4.18
1.0e-003	1.81	0.96	1.81	1.03	1.79	1.27
1.0e-004	1.78	0.93	1.71	0.93	1.73	0.96
1.0e-005	1.83	0.93	1.79	0.95	1.82	0.98

Table 3E

Mean of R^2

τ	T= 75	T= 200	T= 850
1	0.90	0.95	0.99
1.0e-001	0.17	0.26	0.52
1.0e-002	0.03	0.02	0.02
1.0e-003	0.02	0.01	0.00
1.0e-004	0.02	0.01	0.00
1.0e-005	0.02	0.01	0.00

Table 3F

Comparison of MSE's from $\bar{r}$ and $\hat{r}_{T+k|T}$

	T= 75			T= 200			T= 850		
τ	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k|T})$	Ratio
1	0.396	0.055	7.152	0.375	0.050	7.430	0.378	0.051	7.463
1.0e-001	0.019	0.015	1.272	0.009	0.005	1.627	0.005	0.002	3.089
1.0e-002	0.013	0.014	0.977	0.005	0.005	1.003	0.001	0.001	1.035
1.0e-003	0.014	0.014	0.970	0.005	0.005	0.993	0.001	0.001	1.002
1.0e-004	0.014	0.014	0.975	0.005	0.005	0.988	0.001	0.001	0.998
1.0e-005	0.013	0.014	0.977	0.005	0.005	0.989	0.001	0.001	0.998

Notes: The simulated system is: $r_{t+1} = \mu + \beta X_t + \varepsilon_{t+1}$, $X_{t+1} = \mu_x + \phi X_t + u_{t+1}$ var(ε_t)=1, var(u_t)=τ^2 $\times$ var(ε_t). Since we are considering very persistent predictors, we let $\phi = 1$ but similar results obtain for ϕ close to unity. Without loss of generality, we let $\beta = 1$. Note that, given the normalization used in the text, the simulations are invariant with respect to β. τ is a fixed small number. The correlation between u_t and ϵ_t is -0.62, so there is small sample bias (Stambaugh (1999)). The system is simulated 5000 times, for each specification of (τ, T). The first and second moments in Tables A-C are calculated by using sample analogues. Table F compares the out-of-sample forecast of the unconditional vs. the conditional mean of r_t, k periods ahead, where k=0.01 $\times$ T. Very similar results are obtained for fractions other than 0.01.

Table 4A

	The OLS estimator of β in forecasting regression					
	T= 75		T= 200		T= 850	
α	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
0.00	1.00	1.88e-003	1.00	2.67e-004	1.00	1.46e-005
0.20	1.00	1.08e-002	1.00	2.18e-003	1.00	2.18e-004
0.50	1.00	1.40e-001	1.00	5.41e-002	1.00	1.27e-002
0.67	1.00	6.29e-001	1.00	3.32e-001	1.00	1.22e-001
1.00	1.10	1.09e+001	1.00	1.08e+001	1.09	1.14e+001
2.00	1.52	6.43e+004	-9.21	4.18e+005	-47.18	8.05e+006

Table 4B

	The OLS estimator of the root in X_t					
	T= 75		T= 200		T= 850	
α	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
0.00	0.98	1.65e-003	0.99	2.43e-004	1.00	1.38e-005
0.20	0.98	1.68e-003	0.99	2.53e-004	1.00	1.41e-005
0.50	0.98	1.59e-003	0.99	2.51e-004	1.00	1.42e-005
0.67	0.98	1.60e-003	0.99	2.50e-004	1.00	1.40e-005
1.00	0.98	1.67e-003	0.99	2.49e-004	1.00	1.43e-005
2.00	0.98	1.67e-003	0.99	2.31e-004	1.00	1.46e-005

Table 4C

	The OLS estimator of the root in r_t					
	T= 75		T= 200		T= 850	
α	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
0.00	0.82	1.37e-002	0.92	3.10e-003	0.98	2.15e-004
0.20	0.56	3.84e-002	0.71	2.37e-002	0.85	8.75e-003
0.50	0.11	2.21e-002	0.12	1.38e-002	0.13	9.31e-003
0.67	0.02	1.47e-002	0.02	5.76e-003	0.02	1.43e-003
1.00	-0.01	1.32e-002	-0.00	5.03e-003	-0.00	1.19e-003
2.00	-0.01	1.35e-002	-0.00	4.90e-003	-0.00	1.15e-003

Notes: See next page.

Table 4D

Mean of t-stat under Null and Alternative

	T= 75		T= 200		T= 850	
α	null	alt	null	alt	null	alt
0.00	0.06	28.52	-0.03	75.47	-0.02	320.93
0.20	-0.04	11.83	0.02	26.46	0.02	83.42
0.50	0.01	3.26	0.02	5.37	0.04	11.11
0.67	0.00	1.57	0.00	2.17	-0.01	3.48
1.00	0.04	0.41	-0.00	0.38	0.03	0.39
2.00	-0.01	0.00	-0.02	-0.01	-0.06	-0.03

Table 4E

Mean of R^2

α	T= 75	T= 200	T= 850
0.00	0.89	0.95	0.99
0.20	0.61	0.73	0.85
0.50	0.14	0.14	0.13
0.67	0.05	0.03	0.02
1.00	0.02	0.01	0.00
2.00	0.01	0.00	0.00

Table 4F

Comparison of MSE's from $\bar{r}$ and $\hat{r}_{T+k|T}$

	T= 75			T= 200			T= 850		
α	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio
0.00	0.407	0.054	7.519	0.385	0.049	7.779	0.379	0.048	7.903
0.20	0.432	0.114	3.774	0.438	0.088	4.986	0.402	0.067	6.019
0.50	1.378	1.088	1.266	1.341	1.035	1.296	1.376	1.043	1.318
0.67	1.120	1.083	1.034	1.057	1.020	1.036	1.002	0.974	1.029
1.00	1.046	1.069	0.979	0.979	0.985	0.994	0.985	0.986	1.000
2.00	0.994	1.019	0.976	1.004	1.019	0.986	0.956	0.959	0.997

Notes: The simulated system is: $r_{t+1} = \mu + \beta X_t + \varepsilon_{t+1}$, $X_{t+1} = \mu_x + \phi X_t + u_{t+1}$ var(ε_t)=1, var(u_t)=$\tau^2 \times$ var(ε_t). Since we are considering very persistent predictors, we let $\phi = 1$ but similar results obtain for ϕ close to unity. Without loss of generality, we let $\beta = 1$. Note that, given the normalization used in the text, the simulations are invariant with respect to β. $\tau = 1/T^\alpha$, where α is a real positive number. The correlation between u_t and ϵ_t is zero, so there is no bias. The system is simulated 5000 times, for each specification of (α, T). The first and second moments in Tables A-C are calculated by using sample analogues. Table F compares the out-of-sample forecast of the unconditional vs. the conditional mean of r_t, k periods ahead, where k=0.01 $\times$ T. Very similar results are obtained for fractions other than 0.01.

Table 5A

The OLS estimator of β in forecasting regression

	T= 75		T= 200		T= 850	
α	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$	$E(\hat{\beta})$	$Var(\hat{\beta})$
0.00	1.04	2.34e-003	1.02	3.57e-004	1.00	2.01e-005
0.20	1.10	1.38e-002	1.05	3.05e-003	1.02	2.99e-004
0.50	1.38	1.93e-001	1.23	6.76e-002	1.12	1.74e-002
0.67	1.79	7.99e-001	1.58	4.40e-001	1.35	1.56e-001
1.00	4.25	1.39e+001	4.25	1.35e+001	4.38	1.51e+001
2.00	253.34	8.20e+004	669.48	6.19e+005	2890.35	1.05e+007

Table 5B

The OLS estimator of the root in X_t

	T= 75		T= 200		T= 850	
α	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$	$E(\hat{\phi})$	$Var(\hat{\phi})$
0.00	0.98	1.53e-003	0.99	2.48e-004	1.00	1.47e-005
0.20	0.98	1.62e-003	0.99	2.43e-004	1.00	1.35e-005
0.50	0.98	1.82e-003	0.99	2.58e-004	1.00	1.43e-005
0.67	0.98	1.77e-003	0.99	2.54e-004	1.00	1.32e-005
1.00	0.98	1.65e-003	0.99	2.45e-004	1.00	1.47e-005
2.00	0.98	1.76e-003	0.99	2.62e-004	1.00	1.46e-005

Table 5C

The OLS estimator of the root in r_t

	T= 75		T= 200		T= 850	
α	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$	$E(\hat{\phi}_2)$	$Var(\hat{\phi}_2)$
0.00	0.77	1.99e-002	0.90	4.93e-003	0.97	4.03e-004
0.20	0.51	4.17e-002	0.67	2.59e-002	0.83	1.06e-002
0.50	0.10	2.17e-002	0.12	1.30e-002	0.13	8.63e-003
0.67	0.02	1.42e-002	0.02	5.59e-003	0.02	1.39e-003
1.00	-0.01	1.33e-002	-0.01	5.06e-003	-0.00	1.17e-003
2.00	-0.01	1.33e-002	-0.01	4.90e-003	-0.00	1.17e-003

Notes: See next page.

Table 5D

Mean of t-stat under Null and Alternative

	T= 75		T= 200		T= 850	
α	null	alt	null	alt	null	alt
0.00	1.83	29.45	1.74	76.42	1.74	323.67
0.20	1.78	12.85	1.79	27.07	1.77	83.65
0.50	1.77	4.19	1.77	6.25	1.78	11.84
0.67	1.81	2.51	1.77	3.12	1.73	4.44
1.00	1.81	1.31	1.80	1.33	1.80	1.34
2.00	1.89	0.98	1.77	0.95	1.79	0.96

Table 5E

Mean of R^2

α	T= 75	T= 200	T= 850
0.00	0.90	0.95	0.99
0.20	0.65	0.74	0.86
0.50	0.20	0.17	0.15
0.67	0.09	0.05	0.02
1.00	0.03	0.01	0.00
2.00	0.02	0.01	0.00

Table 5F

Comparison of MSE's from $\bar{r}$ and $\hat{r}_{T+k|T}$

		T= 75			T= 200			T= 850	
α	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio	$MSE(\bar{r})$	$MSE(\hat{r}_{T+k\|T})$	Ratio
0.00	0.395	0.055	7.234	0.394	0.051	7.703	0.388	0.053	7.266
0.20	0.469	0.115	4.078	0.435	0.090	4.857	0.404	0.067	5.997
0.50	1.472	1.085	1.357	1.405	1.083	1.297	1.367	1.057	1.294
0.67	1.145	1.060	1.081	1.138	1.073	1.060	1.022	0.975	1.049
1.00	1.020	1.024	0.996	1.044	1.053	0.992	0.991	0.989	1.001
2.00	1.028	1.051	0.978	1.003	1.012	0.991	0.990	0.991	0.999

Notes: The simulated system is: $r_{t+1} = \mu + \beta X_t + \varepsilon_{t+1}$, $X_{t+1} = \mu_x + \phi X_t + u_{t+1}$ var(ε_t)=1, var(u_t)=$\tau^2 \times$ var(ε_t). Since we are considering very persistent predictors, we let $\phi = 1$ but similar results obtain for ϕ close to unity. Without loss of generality, we let $\beta = 1$. Note that, given the normalization used in the text, the simulations are invariant with respect to β. $\tau = 1/T^{\alpha}$, where α is a real positive number. The correlation between u_t and ϵ_t is -0.62, so there is small sample bias (Stambaugh (1999)). The system is simulated 5000 times, for each specification of (α, T). The first and second moments in Tables A-C are calculated by using sample analogues. Table F compares the out-of-sample forecast of the unconditional vs. the conditional mean of r_t, k periods ahead, where k=0.01 $\times$ T. Very similar results are obtained for fractions other than 0.01.

Table 6: Forecasting Error

A: (MSE(Forecastor)/MSE(Mean)-1)*100 from Fixed-Sample Estimation

	1927:1-1998:12	1927:1-1949:12	1950:1-1998:12	1950:1-1979:12	1980:1-1998:12
DY	12.49	-0.32	24.85	-2.24	35.27
TBL	1.63	-0.94	0.67	3.74	-1.54
LONG	1.36	-2.09	0.45	4.97	-3.79
DSPR	0.76	10.43	0.56	-0.08	16.37
TSPR	0.62	-0.05	0.72	8.21	2.04
RTBL	-0.44	6.52	-1.05	0.22	1.82
RR	-0.10	-1.11	0.00	2.93	0.72

B: (MSE(Forecastor)/MSE(Mean)-1)*100 from Rolling Estimation

	1927:1-1998:12	1927:1-1949:12	1950:1-1998:12	1950:1-1979:12	1980:1-1998:12
DY	0.05	-0.54	0.66	-2.26	2.84
TBL	0.30	-0.48	1.94	2.89	-2.01
LONG	1.06	-1.09	1.67	1.76	-2.09
DSPR	0.43	3.73	0.58	0.73	2.15
TSPR	0.10	0.13	1.62	2.99	0.98
RTBL	0.32	3.99	-0.10	0.32	-0.35
RR	0.15	-0.46	0.09	1.56	-0.54

Notes: The entries in Table A represent the percentage *increase* in forecasting error from using a given predictor instead of the unconditional mean. All forecasts are one-period ahead. In each subsample, we estimate the model using as much data as possible, while leaving the last 60 observations for forecasting. Bossaerts and Hillion (1999) use a similar validation procedure. The entries in Table B present the results from a similar comparison, with the exception that the regression is re-estimated at each period before the forecast. Comparing Tables A and B, we reach similar conclusions, thus lending support to our claim that the lack of predictability is not due to a lack of stability in the relation but rather to a lack of signal

Table 7A: Median Unbiased Estimates $\hat{\alpha}^{MU}$ and 99 percent CI's

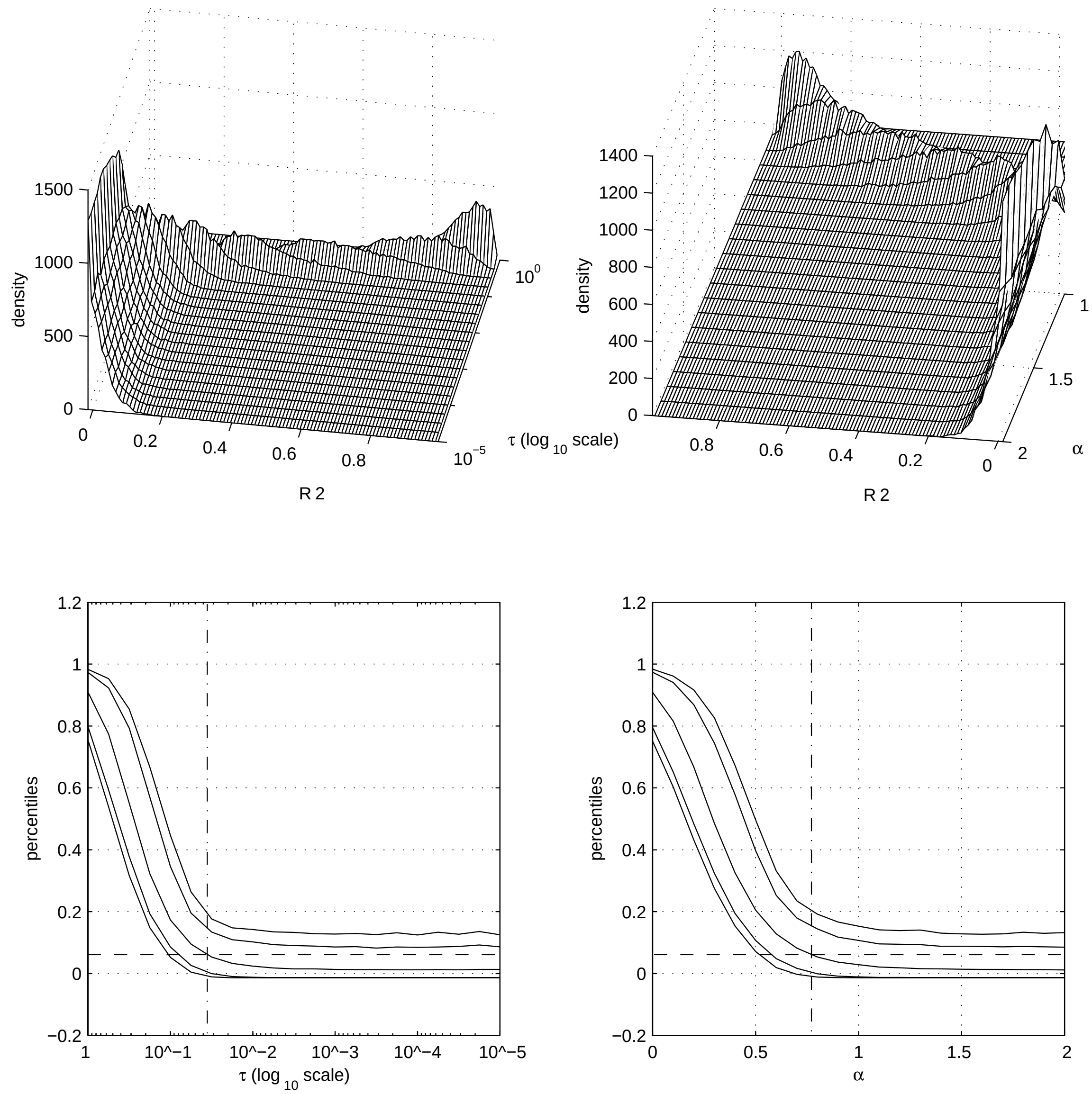
	1927:1-1998:12	1927:1-1949:12	1950:1-1998:12	1950:1-1979:12	1980:1-1998:12
DY	0.59	0.74	0.50	0.32	0.95
	(0.48 – 0.73)	(0.57 – 1.20)	(0.38 – 0.64)	(0.20 – 0.47)	(0.60 – 1.20)
TBL	0.53	0.54	0.40	0.36	0.41
	(0.40 – 0.65)	(0.36 – 0.73)	(0.29 – 0.55)	(0.22 – 0.51)	(0.25 – 0.59)
LONG	0.73	0.61	0.63	0.60	0.56
	(0.56 – 1.20)	(0.44 – 0.86)	(0.47 – 0.81)	(0.45 – 0.83)	(0.39 – 0.76)
DSPR	0.64	0.75	0.52	0.52	0.67
	(0.50 – 0.77)	(0.53 – 1.20)	(0.40 – 0.68)	(0.36 – 0.70)	(0.45 – 1.20)
TSPR	0.43	0.64	0.39	0.36	0.48
	(0.31 – 0.56)	(0.44 – 0.97)	(0.26 – 0.52)	(0.22 – 0.52)	(0.30 – 0.67)
RTBL	0.40	0.66	0.28	0.39	0.27
	(0.29 – 0.54)	(0.45 – 1.20)	(0.16 – 0.43)	(0.25 – 0.54)	(0.12 – 0.46)
RR	0.34	0.36	0.25	0.30	0.26
	(0.22 – 0.49)	(0.20 – 0.51)	(0.13 – 0.39)	(0.17 – 0.46)	(0.11 – 0.43)

Table 7B: Biased, Least Squares Estimates $\hat{\alpha}^{OLS}$

	1927:1-1998:12	1927:1-1949:12	1950:1-1998:12	1950:1-1979:12	1980:1-1998:12
DY	0.72	0.83	0.66	0.49	0.97
TBL	0.68	0.73	0.57	0.54	0.60
LONG	0.88	0.79	0.77	0.77	0.73
DSPR	0.77	0.89	0.69	0.71	0.89
TSPR	0.59	0.82	0.55	0.54	0.67
RTBL	0.56	0.84	0.46	0.57	0.46
RR	0.50	0.55	0.41	0.47	0.45

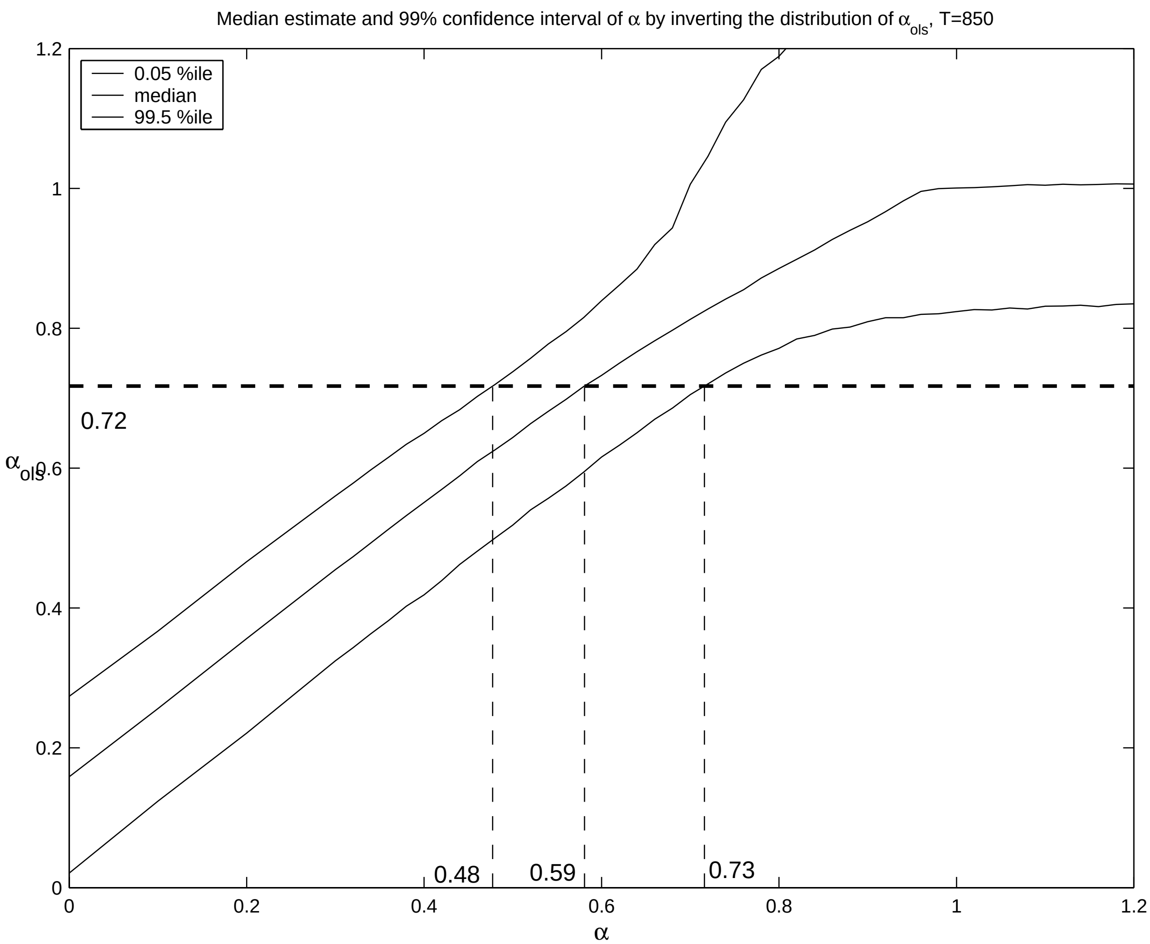
Notes: Table 7A displays median-unbiased estimates of α, the index of signal strength, for different periods and predictors. Values of $\hat{\alpha}^{MU}$ close to zero indicate strong signal. As $\hat{\alpha}^{MU}$ increases, the signal decreases. For α around or higher than 0.5, the signal is weak enough that conditional and unconditional forecasts will produce similar MSE's (see Proposition 5). The estimates $\hat{\alpha}^{MU}$s are computed by inverting the α^{OLS} statistic, as discussed in section 4. The values of the $\hat{\alpha}^{OLS}$ statistic are shown in Table 7B. The results in Table 7A lead us to two conclusions. First, the signal-noise ratio in our predictors is low. The 99 percent confidence intervals of all conditioning variables, in all sub-samples does not contain zero. Moreover, most of the estimates are either insignificantly different from 0.5, or even higher. Second, the estimates $\hat{\alpha}^{MU}$ vary significantly from sample to sample and from predictor to predictor, indicating that it is unlikely that any single predictor will dominate the rest in all periods. Third, the relative rate (RR) must be a good predictor of returns, in all sub-periods. The estimates $\hat{\alpha}^{MU}$ of RR are consistently lower than the rest of the predictors, although the difference is not always significant. Recalling the results from Table 6, this conclusion is borne out by the data. Those estimates are also in agreement with the results in Campbell (1991). The log dividend yield performs particularly poorly in the last sub-period.

Figure 1: Distribution of R^2 as a function of the signal-noise ratio



Notes: Figure 1 displays the distribution of the R^2 statistic as a function of the signal-noise ratio. The top panels show the entire density (as a function of τ or α), whereas the lower panels display percentiles 0.5, 5, 50, 95, 99.5. In the lower figures, the horizontal line marks the value $R^2 = 0.06$, assumed by Bossaerts and Hillion to be the true value of the goodness-of-fit statistic. Note that we find significantly smaller values of R^2 using slightly different data (c.f. Table 1). The vertical dashed-dotted line facilitates the reading of the 95 and 99 percent confidence intervals for the median value of $R^2 = 0.06$.

Figure 2: Estimation of α^{MU} and Centered 99 % Confidence Intervals by diagrammatic inversion of the α^{OLS} statistic of the log dividend yield variable, T=850



Notes: Figure 2 provides a graphic inversion of the distribution of α^{OLS} in order to find α^{MU} and centered 99 % confidence intervals. The distribution is computed by simulating the system (1-5) for various values of α. The 0.5th, 50th (median), and 99.5th quantiles of the distribution of α^{OLS} are plotted in solid lines. The estimated $\hat{\alpha}^{OLS} = 0.72$ in the log dividend yield regression is denoted as a dashed horizontal line. The quantiles $Q_{0.005}(0.72)$, $Q_{0.50}(0.72)$, and $Q_{0.995}(0.72)$ are inverted as shown by the dashed vertical lines. The 99% confidence intervals $Q^{-1}_{0.005}(0.72) = 0.48$ and $Q^{-1}_{0.995}(0.72) = 0.73$, and the median-estimate $\hat{\alpha}^{MU} = Q^{-1}_{0.50}(0.72) = 0.59$ can be read off the horizontal axis.

Appendix A

Calculations for equations (1-2), with τ fixed. We have

$$\begin{aligned} Y_{t+1} &= \mu_y + \beta X_t + \varepsilon_{t+1} \\ X_{t+1} &= X_t + u_{t+1} \\ u_t &= \tau v_t \end{aligned}$$

Then $\hat{\beta} = \left(\sum_{t=1}^T Y_{t+1}\left(X_t - \overline{X}\right)\right)\left(\sum_{t=1}^T \left(X_t - \overline{X}\right)^2\right)^{-1}$

$= \beta + \left(\sum_{t=1}^T \left(X_t - \overline{X}\right)\varepsilon_{t+1}\right)\left(\sum_{t=1}^T \left(X_t - \overline{X}\right)^2\right)^{-1}$

$= \beta + \left(\sum_{t=1}^T \left(\sum_{j=1}^t u_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t u_j\right)\varepsilon_{t+1}\right) \times \left(\sum_{t=1}^T \left(\sum_{j=1}^t u_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t u_j\right)^2\right)^{-1}$

$= \beta + \left(\tau \sum_{t=1}^T \left(\sum_{j=1}^t v_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t v_j\right)\varepsilon_{t+1}\right) \times \left(\tau^2 \sum_{t=1}^T \left(\sum_{j=1}^t v_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t v_j\right)^2\right)^{-1}.$

Then

$$\begin{aligned} T\left(\hat{\beta} - \beta\right) &= \frac{\frac{\tau}{T}\sum_{t=1}^T \left(\sum_{j=1}^t v_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t v_j\right)\varepsilon_{t+1}}{\frac{\tau^2}{T^2}\sum_{t=1}^T \left(\sum_{j=1}^t v_j - \frac{1}{T}\sum_{t=1}^T \sum_{j=1}^t v_j\right)^2} \\ &\Rightarrow \frac{1}{\tau}\frac{\int_0^1 W_2^\mu(s)dW_1(s)}{\int_0^1 \left(W_2^\mu(s)\right)^2 ds} \end{aligned}$$

Define the usual t-statistic:

$$t_{\hat{\beta}} = \frac{\hat{\beta} - 0}{se(\hat{\beta})} = \frac{\left(\hat{\beta} - 0\right)\left(\sum_{t=1}^T \left(X_t - \overline{X}\right)^2\right)^{1/2}}{\left(\frac{1}{T-1}\sum_{t=1}^T \left(Y_{t+1} - \hat{\beta}X_t\right)^2\right)^{1/2}} = \frac{A * B}{C}$$

First, under the null that $\beta = \beta_0$, from above, $T\left(\hat{\beta} - \beta_0\right) \Rightarrow \frac{1}{\tau}\frac{\int_0^1 W_2(s)dW_1(s)}{\int_0^1 W_2^2(s)ds}$. Second $\frac{1}{T}\left(\sum_{t=1}^T X_t^2\right)^{1/2} = \left(\frac{1}{T^2}\sum_{t=1}^T X_t^2\right)^{1/2} \Rightarrow \left(\tau^2 \int_0^1 W_2^2(s)ds\right)^{1/2}$. Third, $C \to_p \left(\sigma_\varepsilon^2\right)^{1/2}$. Putting things together, we have the usual result:

$$t_{\hat{\beta}} \Rightarrow \frac{\int_0^1 W_2^\mu(s)dW_1(s)}{\left(\int_0^1 \left(W_2^\mu(s)\right)^2 ds\right)^{1/2}}$$

Proposition 1 Proof: Suppose $\tau = 1/T^\alpha$. Then, similarly to the previous calculations, $\hat{\beta} = \frac{1}{T^\alpha}\left(\sum_{t=1}^T \left(\sum_{j=1}^t v_j\right)\varepsilon_{t+1}\right)\left(\frac{1}{T^{2\alpha}}\sum_{t=1}^T \left(\sum_{j=1}^t v_j\right)^2\right)^{-1}$

$$T^{(1-\alpha)}\left(\hat{\beta}-\beta\right) = T^{(1-\alpha)}\frac{\frac{1}{T^{\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)^2}$$

$$= \frac{\frac{1}{T^{\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)^2}\frac{\frac{1}{T^{(1-\alpha)}}}{\frac{1}{T^{2(1-\alpha)}}}$$

$$= \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)\varepsilon_{t+1}}{\frac{1}{T^{2}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)^2}$$

$$T^{(1-\alpha)}\left(\hat{\beta}-\beta\right) \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2 ds}$$

$\blacksquare$

Proposition 2 Proof: Under the null, the t-statistic is:

$$t_{\hat{\beta}} = \frac{\left(\hat{\beta}-0\right)\left(\sum_{t=1}^{T}X_t^2\right)^{1/2}}{\left(\frac{1}{T-1}\sum_{t=1}^{T}\left(Y_{t+1}-\hat{\beta}X_t\right)^2\right)^{1/2}} = \frac{A*B}{C}$$

First, from above, $T^{(1-\alpha)}\left(\hat{\beta}-\beta\right) \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2 ds}$. Second, $\frac{1}{T^{1-\alpha}}\left(\sum_{t=1}^{T}\left(X_t-\overline{X}\right)^2\right)^{1/2} = \frac{1}{T^{1-\alpha}}\left(\frac{1}{T^{2\alpha}}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)^2\right)^{1/2} =$
$\left(\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{j=1}^{t}v_j-\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{t}v_j\right)\right)^2\right)^{1/2} \Rightarrow \left(\int_0^1\left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}$, and $C \to_p \left(\sigma_\varepsilon^2\right)^{1/2}$. Finally, as above

$$t_{\hat{\beta}} \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}}$$

$\blacksquare$

Proposition 3 Proof: Let $v_t = \varepsilon_t - \mathrm{Proj}(\varepsilon_t|u_t) = \varepsilon_t - \delta u_t$, where Proj() is the linear projection of ε_t on u_t and $\delta = \frac{\sigma_{\varepsilon u}}{\sqrt{\sigma_u \sigma_\varepsilon}}$. Then, $\frac{1}{(1-\delta^2)^{1/2}}\frac{1}{T^{1/2}}\sum_{i=0}^{t}v_t \Rightarrow W_\perp(s)$, where $W_\perp(s)$ is a standard Wiener process, distributed independently of $W_2(s)$ by construction. We can also write it as $W_\perp(s) = \left(W_1(s) - \delta W_2(s)\right)$. Using exactly the same steps as above, it is straight forward to show that $T^{(1-\alpha)}\left(\hat{\beta}-\beta\right) \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2 ds} = \left(1-\delta^2\right)^{1/2}\frac{\int_0^1 W_2^{\mu}(s)dW_\perp(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2 ds} + \delta\frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\int_0^1\left(W_2^{\mu}(s)\right)^2 ds}$. Similarly $t_{\hat{\beta}} \Rightarrow \frac{\int_0^1 W_2^{\mu}(s)dW_1(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} = \left(1-\delta^2\right)^{1/2}\frac{\int_0^1 W_2^{\mu}(s)dW_\perp(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}} + \delta\frac{\int_0^1 W_2^{\mu}(s)dW_2(s)}{\left(\int_0^1\left(W_2^{\mu}(s)\right)^2 ds\right)^{1/2}}$, and that completes the proof. $\blacksquare$

Proposition 4 Proof: Recall that $R^2 = \frac{\hat{\beta}^2\sum_{t=1}^{T}\left(X_t-\overline{X}\right)^2}{\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2}$. The denominator is $\sum_{t=1}^{T}\left(Y_t-\overline{Y}\right)^2 = \hat{\beta}^2\sum_{t=1}^{T}\left(X_t-\overline{X}\right)^2 + \sum_{t=1}^{T}\varepsilon_t^2 + LOT$, where LOT denotes terms of lower stochastic order for any α. For $\alpha < 1/2$, the first term dominates, for $\alpha > 1/2$ the second term dominates, and for $\alpha = 1/2$ the two terms are of the same O_p order. Then, for $\alpha < 1/2$, $R^2 = \frac{\hat{\beta}^2 \frac{1}{T^{2-2\alpha}} \sum_{t=1}^{T} (X_t - \overline{X})^2}{\hat{\beta}^2 \frac{1}{T^{2-2\alpha}} \sum_{t=1}^{T} (X_t - \overline{X})^2 + o_p(1)} \to^p 1$. In the case $\alpha > 1/2$, $R^2 = \frac{o_p(1)}{op(1) + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2} \to^p 0$. For the borderline case $\alpha = 1/2$, we have $R^2 = \frac{\hat{\beta}^2 \frac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})^2}{\hat{\beta}^2 \frac{1}{T} \sum_{t=1}^{T} (X_t - \overline{X})^2 + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2} \Rightarrow \frac{\beta^2 \int_0^1 (W_2^\mu(s))^2 ds}{\beta^2 \int_0^1 (W_2^\mu(s))^2 ds + \sigma_\varepsilon}$, which completes the proof. ■

Proposition 5 Proof: Before proceeding, note that $Y_{T+k} = \mu_y + \beta X_{T+k-1} + \varepsilon_{T+k} = \mu_y + \beta X_T + \beta \sum_{i=1}^{k-1} u_{T+i} + \varepsilon_{T+k}$, where $u_t = \frac{1}{T^\alpha} v_t$ and $k = [\kappa T]$. Then

$$T^{-(1/2-\alpha)} Y_{T+k} = T^{-(1/2-\alpha)} \left(\mu_y + \beta X_T + \beta \sum_{i=1}^{k-1} u_{T+i} + \varepsilon_{T+k} \right) \Rightarrow \beta W(1) + \beta R \quad , \alpha < 1/2$$

$$Y_{T+k} \Rightarrow \mu_y + \beta W(1) + \beta R + \varepsilon_{T+k} \quad , \alpha = 1/2$$

$$Y_{T+k} \Rightarrow \mu_y + \varepsilon_{T+k} \quad , \alpha > 1/2$$

where $R \sim N\left(0, \kappa \sigma_v^2\right)$ and is independently distributed from ε_{T+k}. Similarly, the mean of Y_t

$$T^{-(1/2-\alpha)} \overline{Y} \Rightarrow \beta \int_0^1 W(s)\, ds \quad , \alpha < 1/2$$

$$\overline{Y} \Rightarrow \mu_y + \beta \int_0^1 W(s)\, ds \quad , \alpha = 1/2$$

$$\overline{Y} \to^p \mu_y \quad , \alpha > 1/2$$

If we use $\hat{Y}_{T+k|T}$,

$$T^{-(1/2-\alpha)} \hat{Y}_{T+k} = T^{-(1/2-\alpha)} \left(\hat{\alpha} + \hat{\beta} X_T \right) \Rightarrow \beta W(1) \quad , \alpha < 1/2$$

$$\hat{Y}_{T+k} \Rightarrow \mu_y + \beta W(1) \quad , \alpha = 1/2$$

$$\hat{Y}_{T+k} \to^p \mu_y \quad , \alpha > 1/2$$

Then, the asymptotic bias from using $\hat{Y}_{T+k|T}$ is $E\left\{ T^{-(1/2-\alpha)} \left(Y_{T+k} - \hat{Y}_{T+k|T} \right) \right\} \to 0$ for $\alpha < 1/2$ and $E\left\{ T \left(Y_{T+k} - \hat{Y}_{T+k|T} \right) \right\} \to 0$ for $\alpha \geq 1/2$. Similarly, the asymptotic bias from using $\overline{Y}$ is $E\left[T^{-(1/2-\alpha)} \left(Y_{T+k} - \overline{Y} \right) \right] \to 0$ for $\alpha < 1/2$ and $E\left[T \left(Y_{T+k} - \overline{Y} \right) \right] \to 0$ for $\alpha \geq 1/2$. Therefore, both forecasts are asymptotically unbiased, for all $a's$.

The expressions for the asymptotic variances can be computed in the same fashion, using the fact that $E\left(\frac{X_T}{T^{1/2}} \frac{(X_1 + ... + X_T)}{T^{3/2}} \right) = \frac{1}{T^2}(EX_1 X_T + ... + X_T^2) = \frac{1}{T^2} E(\varepsilon_1^2 + \left(\varepsilon_1^2 + \varepsilon_2^2\right) + ... + \left(\varepsilon_1^2 + \varepsilon_2^2 + ... + \varepsilon_T^2\right) + \textit{cross terms}) = \frac{1}{T^2}(\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2 + ... + T\sigma_\varepsilon^2) = \frac{\sigma_\varepsilon^2}{T^2} \frac{1}{2} T(T+1) \to \frac{1}{2}\sigma_\varepsilon^2$. Similarly, $E\left(\frac{X_1 + ... + X_T}{T^{3/2}} \right)^2 = \frac{1}{T^3} E(X_1^2 + ... + X_T^2 + 2\sum_{i>j} X_i X_j) = \frac{1}{T^3}\left[\sigma_\varepsilon^2 + 2\sigma_\varepsilon^2 + ... + T\sigma_\varepsilon^2 \right] + 2\sigma_\varepsilon^2 (T + 2(T-1) + 3(T-2) + ... + (T-1)) = \frac{\sigma_\varepsilon^2}{T^3} \left\{ \sum_{i=1}^{T} i + 2\sum_{i=1}^{T} (T-i+1) i \right\} = \frac{\sigma_\varepsilon^2}{T^3} (\frac{T(T+1)}{2} + \frac{1}{6}T^3 + \frac{1}{2}T^2 + \frac{1}{3}T) \to \frac{1}{3}\sigma_\varepsilon^2$. ■

Appendix B

Here we use our small signal-noise asymptotics to demonstrate analytically that spurious regression is unlikely to be a problem in predictive regressions. Ferson, Sarkissian, and Simin (2000) posit the following equation for realized returns:

$$r_{t+1} = \mu + Z_t^* + u_t$$

where Z_t^* are unobservable expected returns and u_t are shocks or unexpected returns. An econometrician tries to forecast returns using the following model

$$r_{t+1} = \mu + \delta Z_t + e_t$$

where Z_t is an observable, persistent predictor, such as the dividend yield, modeled as $Z_t = \phi Z_{t-1} + \varepsilon_t$. We will examine the extreme case $\phi = 1$ but as argued in Appendix C, autoregressive roots close to 1 will not change our argument nor our conclusions. Finally, expected returns as also assumed to follow a persistent process or $Z_t^* = Z_{t-1}^* + \varepsilon_t^*$.

If the processes Z_t and $Z_t^*$are uncorrelated, then Phillips (1986) shows that usual estimation and inference will not hold. In fact, the OLS estimates of δ and its t-statistic will not be consistent, producing spuriously large values. However, another implication of the above system is that realized returns must be serially correlated unless the variance of Z_t^* is much smaller than the variance of u_t,or we must have, for all $t's$:

$$Var\left(u_t\right) >> Var\left(\sum_{i=1}^{t} \varepsilon_i^*\right). \tag{9}$$

If the above inequality is not satisfied, then r_t will behave like an integrated process, an implication that is theoretically unappealing and empirically untenable.

Suppose that $\tau\varepsilon_t = v_t$ and $\tau\varepsilon_t^* = v_t^*$ where τ is a real number close to zero, and the processes v_t and v_t^* have a variance of the same order of magnitude as u_t. Therefore, τ must be decreasing function of T, otherwise (9) will not hold.

Let $\tau = \frac{1}{T^\alpha}$, for $\alpha \geq 0$ (again, otherwise (9) will be violated.) First, note that we can write $Z_t = \tau \sum_{j=1}^t v_j$ and $r_{t+1} = \mu + \tau \sum_{j=1}^t v_j^* + u_{t+1}$. Then, the OLS estimate of δ, after demeaning, is:

$$\begin{aligned} \hat{\delta} &= \frac{\sum_{t=1}^T r_{t+1} Z_t}{\sum_{t=1}^T (Z_t)^2} \\ &= \frac{\sum_{t=1}^T \left(\tau \sum_{j=1}^t v_j^* + u_{t+1}\right)\left(\tau \sum_{j=1}^t v_j\right)}{\sum_{t=1}^T \left(\tau \sum_{j=1}^t v_j\right)^2}. \end{aligned}$$

Therefore, using the FCLT as in Appendix A, we can see that:

$$
\begin{aligned}
T^{\alpha}\hat{\delta} &= \frac{\sum_{t=1}^{T}\frac{1}{T^{\alpha}}\sum_{j=1}^{t}\frac{v_j^*}{T^{1/2}}\left(\sum_{j=1}^{t}\frac{v_j}{T^{1/2}}\right)+\sum_{j=1}^{t}\frac{v_j}{T^{1/2}}\frac{u_{t+1}}{T^{1/2}}}{\sum_{t=1}^{T}\left(\sum_{j=1}^{t}\frac{v_j}{T^{1/2}}\right)^{2}} \\
&= O_p(1).
\end{aligned}
$$

Hence, if $\alpha \geq 0$, $\hat{\delta}$ is a consistent estimator of $\delta$and we do not have a problem of spurious regression.

In summary, if we require realized returns to behave as a stationary (more precisely, I(0)) process then τ must be a decreasing function of T. But if $\tau = 1/T^a$, for $\alpha \geq 0$, then we cannot have a spurious correlation.

Appendix C

Note that the same conclusions will hold whether the forecasting variable X_t has a root at or close to unity. We are interested in forecasters that have a largest autoregressive root ϕ close to unity. The literature has modeled such processes by writing $\phi = 1 + \frac{c}{T}$, or ϕ is in a decreasing (at rate T) neighborhood of 1. The exact unit root case is embedded as $c = 0$. We will show that the arguments for $\phi = 1$ generalize to $\phi = 1 + \frac{c}{T}$.

Assuming $X_0 = 0$, for $\phi = 1$, we have $Var\left(\frac{1}{\sqrt{T}}X_t\right) = \sigma_u^2 \frac{t}{T} = \sigma_u^2 s$. This convergence was needed in all calculation in Appendices A and B. Similarly, if $\phi = 1 + \frac{c}{T}$, $Var\left(\frac{1}{\sqrt{T}}X_t\right) = \sigma_u^2 \frac{\phi^{2t}-1}{\phi^2-1} \to -\frac{\sigma_u^2}{2c}\left[1 - e^{2cs}\right]$, as $T \to \infty$. This is the variance of an Ornstein-Uhlenbeck process, defined by, $dJ_c(s) = cJ_c(s) + dW(s)$, $J(0) = 0$, defined on the space $[0,1]$. Moreover, $\lim_{c\to 0} -\frac{\sigma_u^2}{2c}\left[1 - e^{2cs}\right] = \sigma_u^2 s$, or the unit root (c=0) obtains as a limiting case of the local-to-unity process. Therefore, all of our results can be generalized to a local-to-unity process.

References

1. Andrews, D., 1993, Exactly Median Unbiased Estimation of First Order Autoregressive/Unit Root Models, Econometrica, 61:1, 139-165.

2. Bossaerts, P. and P. Hillion, 1999, Implementing Statistical Criteria to Select Return Forecasting Models: What do we learn?, Review of Financial Studies, 12:2, 405-427.

3. Campbell, J.Y., 1991, A Variance Decomposition for Stock Returns, The Economic Journal, 101, 157-179.

4. Campbell, J and R. Shiller, 1988, The Dividend-Price ratio and expectations of future dividends and discount factors, Review of Financial Studies, 1:3, 195-228.

5. Cavanagh, C., G. Elliott, and J. Stock, 1995, Inference in models with nearly integrated regressors, Econometric Theory, 11, 1131-1147.

6. Cochrane, J., 1988, How big is the random walk in GNP?, Journal of Political Economy, 96, 893-920.

7. Ferson, Wayne, Sarkissian, Sergei, and Timothy Simin, 2000,"Spurious regressions in financial economics?", unpublished manuscript, University of Washington, Seattle.

8. Goetzmann, W. and P. Jorion, 1993, Testing the Predictive Power of Dividend Yields, Journal of Finance 48:2, 663-679.

9. Goyal, A. and I. Welch, 1999, Predicting the Equity Premium, working paper #4-99, Anderson School of Management, UCLA, Los Angeles.

10. Granger, C. and P. Newbold, 1974, Spurious regressions in Econometrics, Journal of Econometrics, 2, 111-120.

11. Hansen, B., 1992, Convergence to stochastic integrals for dependent heterogeneous processes, Econometric Theory, 8, 485-500.

12. Pantula, S., 1991, Asymptotic distributions of unit-root tests when the process in nearly stationary, Journal of Business and Economic Statistics, 9:1, 63-71.

13. Perron, Pierre, 1988, Trends and random walks in macroeconomic time series, Journal of Economic Dynamics and Control, 12, 297-332.

14. Perron, Pierre, 1989, Testing for a random walk: A simulation experiment of power when the sampling interval is varied, in B. Raj, ed: Advances in Econometrics and Modeling (Kluwer Academic Publishers, Inc., Norwell, Mass.).

15. Phillips, Peter, 1986, Understanding Spurious Regressions in Econometrics, Journal of Econometrics, 33, 311-340.

16. Poterba, J. and L. Summers, 1988, Mean Reversion in Stock Prices, Evidence and Implications, Journal of Financial Economics 22, 27-59.

17. Schwert, G.W., 1989, Tests for unit roots: a monte carlo investigation, Journal of Business and Economic Statistics, 7, 147-160.

18. Shiller, R. and P. Perron, Testing the Random Walk Hypothesis: Power versus Frequency of Observations, Economic Letters, 18, 381-386.

19. Stambaugh, Robert, 1999, Predictive Regressions, Journal of Financial Economics, 54, 375-421.

20. Summers, L., 1986, Does the Stock Market Rationally Reflect Fundamental Values, Journal of Finance, 41:3, 591-603.

21. Stock, J., 1991, Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series, Journal of Monetary Economics, 28, 435-459.

22. Stock, J. and M. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61:3, 783-820.