Title
Predictive Regressions Revisited

Authors
Torous, Walter
Yan, Shu

Publication Date
2000-12-01

#2-00

Predictive Regressions Revisited

December 1999

Walter Torous
Anderson Graduate School of Management
University of California, Los Angeles

and

Shu Yan
Anderson Graduate School of Management
University of California, Los Angeles

Finance Working Paper
Sponsored by:

CMS Capital Management Sciences

Predictive Regressions Revisited

Walter Torous and Shu Yan
Anderson School of Management
UCLA
Los Angeles, CA 90095

December 1999

[1]We would like to thank Ross Valkanov for many helpful discussions and suggestions. We, however, remain responsible for any remaining errors.

Abstract

Statistical inference in predictive regressions depends critically on the stochastic properties of the posited explanatory variable, in particular, its order of integration. However, confidence intervals for the largest autoregressive root of explanatory variables commonly used in predictive regressions, including the dividend yield, the book-to-market ratio, and the term and default spreads, confirm uncertainty surrounding these variables' order of integration. Using a local to unity framework we investigate the effects of uncertainty in an explanatory variable's order of integration on inferences drawn in predictive regressions. We find no evidence that dividend yields or book-to-market ratios can predict one period ahead stock returns. In the case of predictive regressions using long horizon returns, statistical inference depends not only on the explanatory variable's order of integration but also on the length of the horizon itself.

1 Introduction

In a predictive regression rates of return are regressed against the lagged values of a stochastic explanatory variable. Examples include, among others, regressing common stock returns against the dividend yield (Fama and French (1988)), regressing bond returns against the spread between long term and short term yields on bonds (Keim and Stambaugh (1986)), and regressing changes in spot exchange rates against the spread between forward and spot exchange rates (Fama (1984)). The predictability of these rates of return using lagged stochastic explanatory variables has been interpreted as evidence of market inefficiency or, alternatively, as evidence of time varying expected returns in financial markets (see Fama (1991) for a review).

Inference in predictive regressions, however, depends critically on the stochastic properties of the posited explanatory variables. While extant research has assumed that the explanatory variables are stationary, dramatically different null distributions of test statistics result when regressors are integrated or, in other words, when regressors follow a random walk (Phillips (1987b)). Unfortunately, in practice it is rarely known whether a particular time series actually has a unit root. By way of example, scaled stock prices, that is, stock prices normalized by dividends or earnings, are often used to predict future stock returns (Lamont (1998)). Yet the debate as to whether stock prices follow a random walk or not dates back to at least Kendall (1953) and continues to this very day. Given this uncertainty surrounding the order of integration of stock prices themselves, it is not surprising that the order of integration of stock prices divided by a smooth accounting variable is also uncertain. Ignoring this uncertainty results in tests of predictive regressions having asymptotic sizes that can far exceed their nominal levels. The prevailing evidence of predictability then reflects the fact that conventional tests simply reject too often.

We couch our statistical analysis of predictive regressions in a local-to-unity framework (for example, Phillips (1987a)). That is, the predictive regression's explanatory variable is assumed to follow an autoregression with a root *near* to unity in the sense that for a given sample size, however arbitrarily large, we are unable to distinguish the assumed stationary specification from the unit root alternative. It is this inability to differentiate between stationary and nonstationary dynamics which captures the uncertainty surrounding an explanatory variable's order of integration. Deviations from the unit root theory are measured by a noncentrality parameter which when its value is *equal* to zero gives a time series with a unit root, but when the parameter's value is *close* to zero gives a nearly integrated time series. The main effect of this specification is to induce noncentrality in the limiting distribution theory. Using this asymptotic theory, Elliott and Stock (1994) and Cavanagh, Elliott and Stock (1995) demonstrate, at least theoretically, that incorrect inference can result when conventional tests are applied to one period ahead predictive regressions with nearly integrated regressors.

In this paper we explicitly incorporate the stochastic properties of the posited explanatory variable in the estimation of predictive regressions. The explanatory variables that we consider - the dividend yield, the book-to-market ratio, the default spread, and the term spread - have been previously used in predictive regressions to forecast stock returns. While there are asymptotic gains to be had by incorporating this additional information, to the extent that there is uncertainty about a particular explanatory variable's order of integration, it becomes important to recognize this uncertainty in the estimation. Using the local-to-unity framework, we find that only the term spread reliably forecasts one period ahead stock returns and then only in the post-1952 sample period. There is no evidence that dividend yields or book-to-market ratios can forecast. We also extend these results to the arguably more important case of predictive regressions using long horizon returns. We follow Richardson and Stock (1989) and use an alternative asymptotic distribution theory for long horizon statistics which recognizes that even though the sample size is large, the number of nonoverlapping observations may in fact be small resulting in conventional large sample approximations performing poorly in practice.

The plan of this paper is as follows. In Section 2 we construct confidence intervals for the largest autoregressive root of each of the sampled explanatory variables and so quantify the uncertainty surrounding their order of integration. In almost every case the results are consistent with this root being very near to unity if not equal to unity. It is important therefore to explore statistical inference in predictive regressions when the stochastic explanatory variable is non-stationary as well as stationary. We do so in Section 3 where we show that if we *know* an explanatory variable's autoregressive root, regardless if it is unity or not, asymptotic statistical gains are to be had if this information is incorporated in the predictive regression's estimation. Of course, it is never known with certainty that an explanatory variable has a unit root or for that matter is stationary. Section 4 then uses the local-to-unity framework to explicitly recognize the effects of this uncertainty on inferences drawn in predictive regressions. We consider both the case of predicting one period ahead returns as well as long horizon returns. Section 5 concludes the paper.

2 Unit Roots in Explanatory Variables

Consider a stochastic explanatory variable x_t, for example, the log dividend yield, which obeys

$$x_t = \rho x_{t-1} + \eta_t, \quad b(L)\eta_t = \epsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where $b(L) = \sum_{j=0}^{k} b_j L^j$, $b_0 = 1$, L is the lag operator and ϵ_t is a martingale difference sequence. Here we distinguish between x_t's largest autoregressive root, ρ, and the assumed fixed stable roots of $b(L)$ describing x_t's short-run dynamics. This specification of η_t allows for the possibility of serial correlation or heteroskedasticity in the disturbances. We wish to test for a unit root in x_t, $H_0 : \rho = 1$.

To do so, define $\nu_t = (1 - \rho L)^{-1}\eta_t$ and rewrite (1) as

$$x_t = \nu_t, \quad a(L)\nu_t = \epsilon_t, \quad \text{where} \quad a(L) = b(L)(1 - \rho L)$$

which can be rearranged to give

$$x_t = \alpha(1)x_{t-1} + \sum_{j=1}^{k} \alpha^*_{j-1}\Delta x_{t-j} + \epsilon_t \tag{2}$$

where $\alpha(L) = L^{-1}(1 - a(L))$ so $\alpha(1) = 1 - b(1)(1 - \rho)$ and $\alpha^*_i = -\sum_{j=i+1}^{k} \alpha_j$. The augmented Dickey-Fuller (ADF) statistic is the t-statistic testing $\alpha(1) = 1$ in (2), or equivalently, testing $\rho = 1$.[1]

As emphasized by Stock (1991), simply giving point estimates of ρ or reporting the results of unit root tests does not convey the uncertainty surrounding ρ. Confidence intervals for ρ provide a more useful summary measure of a stochastic explanatory variable's persistence by indicating the range of ρ values that are consistent with the observed data. However, the usual approach of constructing asymptotic confidence intervals as the point estimate $\pm$ 2 standard errors is not appropriate in this case because the distribution of the ADF statistic is nonnormal. Additionally, traditional first-order asymptotic theory does not provide a suitable framework for the construction of these confidence intervals because this theory is discontinuous at $\rho = 1$.

Instead, Stock tabulates confidence intervals for ρ constructed using local-to-unity asymptotic theory where the true value of ρ is modeled as being in a decreasing neighborhood of one. Specifically $\rho = 1 + c/T$ where c is a fixed constant and T is the sample size.[2] Under this specification, the asymptotic distribution of the ADF statistic depends only on c and is continuous in c. A confidence interval for c, and so for ρ given T, can then be constructed by inverting the appropriate acceptance region of the ADF statistic. For further details see Stock (1991).

[1] Alternatively, Phillips and Perron (1988) estimate (1) and adjust the test statistic to take account of serial correlation and potential heteroskedasticity in the disturbances η_t.

[2] As noted by Stock, nesting ρ as a function of the sample size is analogous to the usual approach used to study the asymptotic power of statistical tests against local alternatives except that the alternative is in a $1/T$ neighborhood of the null value of unity.

2.1 Data

A variety of stochastic explanatory variables have previously been used in predictive regressions to forecast stock returns. We construct confidence intervals for the largest autoregressive root of a number of these time series variables.

Dividend Yield

Real dividend yields are constructed from monthly returns, with (r) and without (r_0) dividends, of the Center for Research in Security Prices ($CRSP$) value-weighted (VW) market portfolio. Assuming a one dollar investment in the VW portfolio at the end of December 1925, $P(0) = 1$, the value of the portfolio at the end of month t, $P(t)$, is constructed according to $P(t) = (1 + r_0(t))P(t-1)$. Deflating $P(t)$ by the prevailing Consumer Price Index (CPI) gives the real value of the VW portfolio at the end of month t. Dividends on the portfolio in month t are given by $(r(t) - r_0(t))P(t-1)$ and by deflating these dividends by the CPI we obtain corresponding monthly real dividend payments. The real dividend yield in month t is computed by summing the real monthly dividends for the year preceding month t and dividing by the real value of the VW portfolio at t.[3] We investigate the stochastic properties of the logarithm of this real dividend yield series.

DJIA Book-to-Market Ratio

We follow Pontiff and Schall (1998) and use an aggregate book-to-market ratio based on the Dow Jones Industrial Average (DJIA).[4] The DJIA is an index of the stock prices of thirty large U.S. corporations selected to represent a cross section of U.S. industry. The December year-end book value of the DJIA is available from 1920 to 1993. A monthly book-to-market ratio is constructed by dividing the most recent year's book value by the prevailing level of the DJIA. However, to ensure the use of information that was available to market participants, Pontiff and Schall do not incorporate the previous year's book value in their calculations until March of the subsequent year. We investigate the stochastic properties of the logarithm of Pontiff and Schall's DJIA book-to-market ratio.

Default Spread

The default spread is computed monthly and is measured by the logarithm of the difference between the average annualized yield of bonds rated Baa by Moody's and the average annualized yield of bonds rated Aaa by Moody's. These monthly yields are averages of daily data within the month and are obtained from the Federal Reserve's H.15 statistical release.

[3]We follow Fama and French (1988) and do not reinvest monthly dividends. Alternatively, annual dividends can be constructed by assuming monthly dividends are reinvested at prevailing monthly Treasury bill rates (Hodrick (1992)) or reinvested in the VW portfolio itself (Cochrane (1991)). Our conclusions are not affected by these alternative definitions.

[4]We thank Jeff Pontiff for providing us their book-to-market series.

Term Spread

The term spread is measured by the difference between long term and short term yields on Treasury securities.[5] Specifically, the long term yield is taken to be the annualized yield on a Treasury bond with maturity closest to ten years measured at month end and is obtained from the *CRSP* bond file.[6] The short term yield is given by the annualized three month Treasury bill yield obtained from *CRSP's* Fama-Bliss file.

2.2 Results

For each of our sampled stochastic explanatory variables Table I presents 95% confidence intervals for ρ.[7] We provide results using the entire time series of data and, to investigate the robustness of our conclusions, using the pre-1952 and post-1952 subsamples. In almost every case these 95% confidence intervals include the unit root $\rho = 1$. The exceptions include the *log* dividend yield series over the 1926:12 to 1994:12 sample period whose upper limit of 0.996 is nearly indistinguishable from one. While the 95% confidence interval for the term spread series based on the entire sample period does not contain one, the interval based on the pre-1952 subsample does. *OLS* estimates of ρ based on an $AR(1)$ autoregression, $\hat{\rho}$, are also tabulated in Table I and in almost every case are seen to lie within the corresponding 95% confidence interval.[8]

Confidence intervals for ρ based on the entire time series of observations tend to be narrower than those based on the subsamples with their fewer number of observations. However, the fact remains that even confidence intervals based on the pre-1952 sample period, which contain the fewest number of observations, are still relatively narrow about unity. For example, the 95% confidence interval for ρ using the *log* default spread series over the 1926:12 to 1951:12 sample period is $(0.984, 1.015)$. The results of Table I are consistent with the measure of persistence ρ for stochastic explanatory variables typically used in predictive regressions being very near to unity if not actually equal to unity.

[5]Unlike the default spread, the term spread can be negative and so we do not use a logarithmic transform of this series.

[6]We thank Bruno Gerard for making this data available to us.

[7]We follow Nelson and Plosser (1982) to determine the maximum lag length k.

[8]Recall that the $AR(1)$ autoregression omits higher order autoregressive terms while the confidence intervals are based on ADF statistics which include these higher order terms. As a result, the reported $\hat{\rho}$ need not lie in ρ's 95% confidence interval. By way of example, if the autoregression using the term spread series over the entire sample period is augmented with $k = 6$ higher order autoregressive terms, the resultant OLS ρ estimate of 0.956 is seen to lie in the corresponding 95% confidence interval.

3 Predictive Regressions

Confidence intervals for ρ based on data for a number of stochastic explanatory variables previously used in predictive regressions typically contain the unit root. However, these intervals also include values of ρ different from one and which have substantially different implications for the stochastic behavior of the explanatory variables. It is important, therefore, to explore statistical inference in predictive regressions when the stochastic explanatory variable is non-stationary as well as stationary.

3.1 Small Sample Bias in Predictive Regressions with a Stationary Explanatory Variable

Consider the following predictive regression in which a variable, y_t, is regressed against the lagged value of a stochastic explanatory variable, x_{t-1}:

$$y_t = \alpha + \beta x_{t-1} + u_t, \ t = 1, \ldots, T. \tag{3}$$

In this section we assume that the explanatory variable is *known* to be stationary but persistent:

$$x_t = \mu + \rho x_{t-1} + v_t, \ t = 1, \ldots, T. \tag{4}$$

Stationarity requires that $\rho < 1$ and persistence follows if $\rho \approx 1$.

To complete this regression framework we assume that the disturbance terms u_t and v_t are each independently and identically normally distributed. However, $cov(u_t, v_t) \equiv \sigma_{uv}$ need not equal zero; for example, innovations in stock returns can be expected to be negatively correlated with innovations in the log dividend yield since the stock price enters into each of these variables.

As shown by Stambaugh (1986), under these assumptions the slope of the predictive regression (1), β, will be estimated with bias in small samples. This bias is approximated by $E\{\hat{\beta} - \beta\} \approx -(\sigma_{uv}/\sigma_v^2)T^{-1}(1 + 3\rho)$. While the bias depends on the assumed known value of ρ (as well as σ_{uv} and σ_v^2), in practice ρ is not known. As a result, the bias adjustment is typically implemented using the estimate of ρ obtained by *separately* applying OLS to the autoregression (4).

To understand the nature of this bias in $\hat{\beta}$, it is instructive to first consider the estimation of the autoregression (4). In fact, $\hat{\beta}$'s bias stems from the fact that the slope of the autoregression, ρ, is estimated with bias in small samples. In particular, the OLS estimator $\hat{\rho}$

satisfies

$$\hat{\rho} = \rho + \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x}_{(-1)})(v_t - \bar{v})}{\sum_{t=1}^{T}(x_{t-1} - \bar{x}_{(-1)})^2}$$

where $\bar{x}_{(-1)} = T^{-1}\sum_{t=1}^{T} x_{t-1}$ and $\bar{v} = T^{-1}\sum_{t=1}^{T} v_t$. Now the expectation of the numerator

$$E\sum_{t=1}^{T}(x_{t-1} - \bar{x}_{(-1)})(v_t - \bar{v}) = -\sum_{t=1}^{T} E\bar{v}x_{t-1}$$

is clearly nonzero since $\bar{v}$ contains terms which are correlated with each x_{t-1}, *i.e.*, $v_1, \dots, v_{t-1}$. Although the expectation of a ratio is not the ratio of the expectations of the numerator and denominator, this at least suggests the reasons for this bias. The bias approximated from a first order Taylor series [9] is $E\{\hat{\rho} - \rho\} \approx -T^{-1}(1 + 3\rho)$; see Mariott and Pope (1954) and Kendall (1954).[10] The *OLS* estimator $\hat{\rho}$, however, retains desirable asymptotic properties being consistent and hence asymptotically unbiased.

The small sample bias of the *OLS* estimator $\hat{\beta}$ now follows. To the extent that the disturbances u_t and v_t are correlated, if x_{t-1} is correlated with $v_1, \dots, v_{t-1}$ then x_{t-1} will be correlated with $u_1, \dots, u_{t-1}$. This small sample bias in $\hat{\beta}$ clearly depends on the assumed known value of ρ: if $\rho = 0$ so that x_t is *i.i.d.* then x_{t-1} will not be correlated with past v_ts meaning that even if $\sigma_{uv} \neq 0$ it will be the case that x_{t-1} will not be correlated with past u_ts. The bias also depends on σ_{uv}: if $\sigma_{uv} = 0$ then even if x_t is persistent so that x_{t-1} is correlated with past v_ts it follows that x_{t-1} will not be correlated with past u_ts.

Notice that to this point the time series properties of x_t are *only* used insofar as under the assumption of stationarity the resultant bias of $\hat{\rho}$ imparts bias to the predictive regression's estimated slope coefficient $\hat{\beta}$. Apart from approximating the bias adjustment in $\hat{\beta}$, knowledge of x_t's stochastic properties is not used in estimating the predictive regression.

However, as noted by Nelson and Kim (1993) and others, $\hat{\beta}$'s bias *per se* can affect the inference drawn in predictive regressions. Their intuition is that as $\hat{\beta}$ directly enters the corresponding t-statistic's numerator, a biased value of $\hat{\beta}$ will give an erroneous t-statistic. To correct this, the t-statistic in predictive regressions must be recalculated taking into account the correlation prevailing between the disturbances which gives rise to the bias in $\hat{\beta}$. Nelson and Kim's randomization procedure explicitly takes this correlation into account and so allows bias-corrected inference of predictive regressions.

To investigate the nature of this bias-corrected inference, we conduct the following Monte Carlo study of Nelson and Kim's randomization procedure. To begin with, for each of

[9] As noted by Kendall this approximation is of "doubtful validity for ρ near unity" (page 404) as the required Taylor series expansion may not converge sufficiently quickly.

[10] Table I also presents bias-adjusted estimates of ρ given by $\hat{\rho}_{adj} = \frac{T\hat{\rho}+1}{T-3}$.

$\rho = \{0.90, 0.95, 0.99\}$, $\rho_{uv} = \frac{\sigma_{uv}}{\sigma_u \sigma_v} \equiv \delta = \{-0.5, -0.9\}$, $\sigma_v = \{0.1, 1.0\}$, and $\sigma_u = 1.0$, we generate, under $H_0 : \beta = 0$, time series of length $T = 100$ of x_t and y_t[11] according to expressions (3) and (4), respectively. We then calculate $\hat{\beta}$ as well as $\hat{\rho}$. We repeat this experiment 5000 times for each combination of assumed ρ, δ, σ_v, and σ_u values. For these parameter combinations we also retain the resultant residuals $\{\hat{u}_t\}_{t=1}^{100}$ and $\{\hat{v}_t\}_{t=1}^{100}$ for each of the 5000 replications.

Inference in the absence of the bias correction is based on the sampling distribution of $\hat{\beta}$ as we do not explicitly take into account the correlation between residuals in the calculation of $\hat{\beta}$. In Table II for each combination of assumed ρ, δ, σ_v, and σ_u values the sampling distribution of $\hat{\beta}$ is summarized by its median, as well as its corresponding 10^{th}, and 90^{th} percentiles.

The median value of $\hat{\beta}$ corresponds well to β_{adj} the bias-adjusted value of β also tabulated in Table II which can be calculated directly from Stambaugh's approximation for $\beta = 0$ given the assumed values of ρ, δ, σ_u, σ_v and T.[12] As expected, for fixed T the small sample bias in $\hat{\beta}$ tends to increase as ρ increases, as the absolute value of δ increases, and as σ_v decreases.

Following Nelson and Kim, the effect of the small sample bias in $\hat{\beta}$ on inference can be accounted for by explicitly incorporating the correlation prevailing between the disturbances when calculating the significance of the predictive regression's slope coefficient, now denoted by $\check{\beta}$. To do so, for a given combination of ρ, δ, σ_v, and σ_u values, we take the residuals $\{\hat{u}_t\}_{t=1}^{100}$ and $\{\hat{v}_t\}_{t=1}^{100}$ associated with each of the 10^{th}, 50^{th}, and 90^{th} percentiles of the corresponding $\hat{\beta}$ sampling distribution and randomize each *grouping* of these residuals. Given these randomized residuals we can once again generate, under $H_0 : \beta = 0$, time series of length $T = 100$ of x_t and y_t according to expressions (3) and (4), respectively, and then calculate $\check{\beta}$. This randomization is repeated 500 times to obtain $\check{\beta}$'s sampling distribution whose 10^{th}, 50^{th}, and 90^{th} percentiles are also tabulated in Table II. This sampling distribution can be compared to the sampling distribution of $\hat{\beta}$ to assess the impact of bias-corrected inference.

The sampling distribution of $\check{\beta}$ at the median value of $\hat{\beta}$ ($\hat{\beta}_{50\%}$) differs slightly from the sampling distribution of $\hat{\beta}$ across all of the assumed parameter values. However, $\check{\beta}$'s sampling distribution assuming $\hat{\beta}$ values at their 10^{th} ($\hat{\beta}_{10\%}$) and 90^{th} ($\hat{\beta}_{90\%}$) percentiles depends on the parameter values assumed, especially the assumed value of ρ. For $\rho = 0.90$, the sampling distribution of $\check{\beta}$ at $\hat{\beta}_{10\%}$ is shifted left relative to $\hat{\beta}$'s sampling distribution by approximately the magnitude of the bias in estimating $\hat{\beta}$ while the sampling distribution of $\check{\beta}$ at $\hat{\beta}_{90\%}$ is shifted right by this amount. By contrast, for $\rho = 0.99$ the sampling distribution of $\check{\beta}$

[11]Like Nelson and Kim, we draw x_0 from a normal distribution with mean and variance implied by the corresponding $AR(1)$ representation.

[12]These should not match each other exactly as the sampling distribution of $\hat{\beta}$ is skewed to the right and as a result mean values will exceed median values.

appears to widen as we increase the assumed value of $\hat{\beta}$ from $\hat{\beta}_{10\%}$ to $\hat{\beta}_{90\%}$. However, even at $\hat{\beta}_{90\%}$, the rightmost tail of $\check{\beta}$'s sampling distribution is shifted right by approximately the magnitude of the bias in estimating $\hat{\beta}$.

We conclude from the results of Table II that once the small sample bias in estimating $\hat{\beta}$ is accounted for, slightly more extreme empirical cutoff values are needed to reject $H_0 : \beta = 0$. However, the impact on these cutoff values is at most the magnitude of the bias itself. As the sample size T increases, this bias diminishes as will its effect on inference in predictive regressions.

While the small sample bias in estimating the predictive regression's slope coefficient appears to have but a minimal effect on inference in sample sizes typically encountered in practice, it should be emphasized that this analysis assumes that ρ is fixed at less than one. There is no uncertainty about ρ nor do we recognize the possibility of a unit root in the dynamics of the stochastic explanatory variable. We relax this latter constraint in the next section.

3.2 Predictive Regressions with a Non-Stationary Explanatory Variable

The preceding analysis assumes that the explanatory variable x_t in a predictive regression is stationary. Assume now that it is *known* that the explanatory variable, x_t, has a unit root, $\rho = 1$, and so consider the following:

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t, \qquad t = 1, \ldots, T \qquad (5) \\ \Delta x_t &= \mu + v_t, \qquad\qquad\quad t = 1, \ldots, T. \qquad (6) \end{aligned}$$

Notice that we have explicitly imposed the condition that $\rho = 1$. If $(u_t, v_t) \sim iid\ \mathrm{BVN}(0, \Sigma)$, or equivalently, $(y_t - \alpha - \beta x_{t-1}, \Delta x_t - \mu) \sim iid\ \mathrm{BVN}(0, \Sigma)$, then noting that a joint probability density function can be expressed as a product of conditional and marginal probability density functions we can write the corresponding log-likelihood function as

$$ln\mathcal{L}(\alpha, \beta, \mu, \Sigma) =$$

$$-(T/2)ln(\sigma_u^2 - \sigma_{uv}^2/\sigma_v^2) - (1/2)\sum_1^T (y_t - \alpha - \beta x_{t-1} - (\sigma_{uv}/\sigma_v^2)(\Delta x_t - \mu)^2)/(\sigma_u^2 - \sigma_{uv}^2/\sigma_v^2)$$

$$-(T/2)ln(\sigma_v^2) - (1/2)\sum_1^T (\Delta x_t - \mu)^2/\sigma_v^2.$$

By inspection, the maximum likelihood estimate of β is equivalent to the *OLS* estimate of β obtained from the linear model

$$y_t = \alpha + \beta x_{t-1} + \gamma \Delta x_t + w_t \qquad (7)$$

where $w_t = u_t - (\sigma_{uv}/\sigma_v^2)v_t$. For further details see Phillips (1991).

Phillips's result demonstrates that if a unit root is known to be present then this information should be directly incorporated into the predictive regression's empirical specification. To simply regress y_t on x_{t-1} and not impose the condition that the explanatory variable has a unit root will result in incorrect statistical inference. This follows from the fact that if an explanatory variable is an autoregressive process, the distribution of the corresponding t-statistic depends on the roots of the process and on the correlation properties of the error processes. In our case, if the autoregressive process defining the explanatory variable has a unit root and if the error in the autoregressive process is correlated with the regression error, then the usual t-statistic on β has a nonstandard limiting distribution.[13] Directly imposing the condition that the explanatory variable has a unit root, as in expression (7), produces an efficient estimate of β and the corresponding t-statistic converges asymptotically in distribution to a $N(0,1)$ random variable.

3.3 Predictive Regressions with a Stationary Explanatory Variable

To clarify and further these estimation results, it can be recognized that the predictive regression, expression (5), and the autoregression, expression (6), comprise a seemingly unrelated regression (SUR). That is, the equations are related through their respective disturbances. Of course, if the disturbances are uncorrelated it is optimal to apply OLS equation by equation. Otherwise, it is optimal to estimate the system as a whole. Phillips result demonstrates that if the explanatory variable is known to have a unit root then in this case estimating the system as a whole by SUR reduces to running the multiple regression given by expression (7).

In fact, as the following Lemma demonstrates, these conclusions continue to hold regardless of whether the explanatory variable has a unit root ($\rho = 1$), as assumed by Phillips, or is stationary ($\rho = \rho_0 < 1$) so long as the parameters of the autoregression are assumed to be *known*.

To see this consider the following SUR model

$$\begin{aligned} y_{1t} &= x_{1t}'\beta_1 + \epsilon_{1t} \\ y_{2t} &= x_{2t}'\beta_2 + \epsilon_{2t} \end{aligned}$$

[13]In particular, from Fuller's (1996) Theorem 10.3.2 the t-statistic on β converges asymptotically in distribution to $\delta\hat{\tau} + (1-\delta^2)^{\frac{1}{2}}Z$ where δ denotes the correlation between u_t and v_t, Z is a $N(0,1)$ random variable and $\hat{\tau} = [s^{-2}\sum_{t=2}^{T} x_{t-1}^2]^{\frac{1}{2}}(\hat{\rho}-1)$ for $s^2 = (T-2)^{-1}\sum_{t=2}^{T}(x_t - \hat{\rho}x_{t-1})^2$.

$$\vdots$$

$$y_{nt} = x'_{nt}\beta_k + \epsilon_{nt}$$

where we assume the residuals $e_t = (\epsilon_{1t} \cdots \epsilon_{nt})' \sim iid\ N(0, \Omega)$. Here x_{it}, $1 \leq i \leq n$, is a $(k_i \times 1)$ vector of explanatory variables for the ith-equation, and β_i is a $(k_i \times 1)$ vector of coefficients. Suppose the variables y_{it}, $1 \leq i \leq n$, are categorized into two groups as represented by the vectors $Y_{1t} = (y_{1t} \cdots y_{mt})'$ and $Y_{2t} = (y_{(m+1)t} \cdots y_{nt})'$ with corresponding coefficient vectors: $B_1 = (\beta'_1 \cdots \beta'_m)'$ and $B_2 = (\beta'_{m+1} \cdots \beta'_n)'$. Then this system of equations can be written in a compact form as:

$$\begin{aligned} Y_{1t} &= X'_{1t}B_1 + e_{1t} \\ Y_{2t} &= X'_{2t}B_2 + e_{2t}, \end{aligned} \tag{8}$$

where X'_{1t} and X'_{2t} are the following matrices:

$$X'_{1t} = \begin{pmatrix} x'_{1t} & 0 & \cdots & 0 \\ 0 & x'_{2t} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x'_{mt} \end{pmatrix}, X'_{2t} = \begin{pmatrix} x'_{(m+1)t} & 0 & \cdots & 0 \\ 0 & x'_{(m+2)t} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & x'_{nt} \end{pmatrix}.$$

Lemma 1 Consider the SUR model (8). Suppose that the coefficient vector B_2 is known and only B_1 is to be estimated. Then the maximum likelihood estimate of B_1 is given by the OLS estimate of the coefficient of X'_{1t} in the following multiple regression:

$$Y_{1t} = X'_{1t}B_1 + C(Y_{2t} - X'_{2t}B_2) + e_{1 \cdot 2t},$$

where $C = \Omega_{12}\Omega_{22}^{-1}$ and $e_{1 \cdot 2t} = e_{1t} - \Omega_{12}\Omega_{22}^{-1}e_{2t}$.

Proof: See the Appendix.

If the stochastic explanatory variable x_t is known to have a unit root, $\rho = 1$, then y_t should be regressed against x_{t-1} *and* the first-difference of x_t, $\Delta x_t \equiv x_t - x_{t-1}$ to ensure that the usual regression t-statistic correctly assesses the significance of the lagged explanatory variable. Alternatively, if x_t is known to be stationary and $\rho = \rho_0 < 1$, the Lemma gives that y_t should be regressed against x_{t-1} *and* the quasi-difference of x_t, $x_t - \rho_0 x_{t-1}$:

$$y_t = \alpha + \beta x_{t-1} + \gamma(x_t - \rho_0 x_{t-1}) + w_t \tag{9}$$

to ensure that the usual regression t-statistic correctly assesses the significance of the lagged explanatory variable.

If researchers are prepared to treat the autoregression's slope coefficient ρ as known, then optimal statistical inference in a predictive regression requires that this knowledge of ρ be explicitly incorporated in the estimation. The correlation of the disturbances not only implies that the predictive regression's slope parameter will be estimated with bias in small samples but, more importantly, also implies that asymptotic statistical gains are to be had if the system *as a whole* is estimated, or by the Lemma, equivalently, if this knowledge of ρ is explicitly incorporated in the predictive regression equation.

3.4 Results

To illustrate these effects on inference in predictive regressions, Table III provides the results of estimating predictive regressions for each of our sampled series but where we vary the information incorporated about the posited explanatory variable's autoregressive process.

For comparison purposes, we first present the resultant t-statistics when *log* one month real returns to the *VW* portfolio are regressed against lagged values of the explanatory variables *without* incorporating any information about the explanatory variable's autoregressive process. This corresponds to the standard implementation of a predictive regression. We also tabulate the estimated correlation coefficient $\hat{\delta}$ between the estimated disturbances of the predictive regression, $\hat{u}_t$, and the autoregression, $\hat{v}_t$. The next two columns give these estimation results after incorporating the assumption that the explanatory variable's autoregressive process is stationary and characterized by its historical estimate, $\hat{\rho}$, and its bias-adjusted historical estimate, $\hat{\rho}_{adj}$, respectively. The final column assumes that the explanatory variable has a unit root, $\rho = 1$. To ensure comparability across these cases, t-statistics based on Newey-West standard errors corrected for heteroskedasticity and serial correlation are presented throughout.

Consistent with the results of Campbell, Lo, and MacKinlay (1997), the standard predictive regression of *log* real returns against lagged *log* real dividend yields shows a significant predictive relation over the post-1952 sample period ($t = 2.99$). However if we impose the condition that *log* real dividend yields follow a random walk, $\rho = 1$ then the significance of this post-1952 evidence dramatically decreases ($t = 1.87$). On the other hand, if it is assumed that *log* real dividend yields are stationary and $\rho = \hat{\rho}$, the post-1952 evidence becomes extremely significant ($t = 11.75$), so much so that under this assumption a significant predictive relation prevails over the entire sample period as well ($t = 2.69$).

The standard predictive regression using the lagged *log* default spread provides no reliable evidence of being able to forecast one month ahead *log* real returns over either the entire sample or the subsamples. This conclusion continues to hold regardless of whether we impose the condition that the *log* default spread is stationary or not.

Using the lagged *log* book-to-market ratio as an explanatory variable, the standard predictive regression shows some mild evidence ($t = 1.65$) of predictability over the entire sample period but consistent with the results of Pontiff and Schall (1998), the evidence is weakest in the latter subsample period ($t = 0.90$). However, if we impose the condition that the *log* book-to-market ratio is stationary then this evidence over the entire sample period is enhanced (for example, $t = 3.62$ for $\rho = \hat{\rho}$), while if we assume that $\rho = 1$ this evidence of predictability vanishes ($t = 0.34$).

In contrast, the lagged term spread reliably forecasts one month ahead *log* real returns in the post-1952 sample period across all the regression specifications. However, there is no evidence of predictability in the pre-1952 sample period as well as the entire sample period, and once again these conclusions hold across all the regression specifications.

The clear message of Table III is that incorporating information about the explanatory variable's autoregressive process can have a potentially significant effect on inferences drawn in predictive regressions, especially when the disturbance terms to the predictive regression and the autoregressive process are highly correlated. Furthermore, the particular specification of the autoregressive process assumed, for example, the explanatory variable being stationary versus having a unit root, may also have significant implications.

Of course, it is never known with certainty that an explanatory variable has a unit root, or for that matter, is stationary. It is important then that we recognize the effects of uncertainty in ρ on inferences drawn in predictive regressions.

4 Predictive Regressions when the Explanatory Variable's Order of Integration is Unknown

As evidenced by the results of Table I, very often we are uncertain as to whether a stochastic explanatory variable, x_t, is stationary, $\rho < 1$, or is non-stationary, $\rho = 1$. To capture this uncertainty in ρ, we nest ρ in a $1/T$ neighborhood of one, $\rho = 1 + c/T$ where c is a fixed constant and T is the sample size. Nesting ρ in a $1/T$ neighborhood of one captures uncertainty in ρ in the sense that for a given sample size we are unable to distinguish this stationary specification from the unit root alternative, $\rho = 1$.

In this section we explore statistical inference in predictive regressions within this local to unity framework. We first investigate the consequences of using near integrated regressors to predict, as assumed up to now, one period ahead returns and then to predict longer horizon returns.

4.1 One Period Returns

Elliott and Stock (1994) and Cavanagh, Elliott and Stock (1995) demonstrate that the use of near integrated regressors introduces potentially substantial size distortions in conventional tests of time series regressions, both in small samples and asymptotically.[14] This suggests, at least theoretically, that the extant evidence of predictability may simply reflect the fact that these conventional statistics reject too often.

Suppose that one period *log* real returns, y_t, are regressed against lagged values of a stochastic explanatory variable, x_{t-1}:

$$y_t = \alpha + \beta x_{t-1} + u_t, \;\; t = 1, \ldots, T \tag{10}$$

and the dynamics of x_t are described by

$$x_t = \mu + \rho x_{t-1} + \sum_{i=1}^{k} \zeta_i \Delta x_{t-i} + \nu_t, \;\; t = 1, \ldots, T \tag{11}$$

where $\epsilon_t = (u_t, \nu_t)'$ is a martingale difference sequence, $E(\epsilon_t \epsilon_t' \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = \Sigma$ with typical element $\sigma_{u\nu}$. As before, let δ denote the correlation between the disturbance terms and now nest ρ in a $1/T$ neighborhood of one, $\rho = 1 + c/T$ where c is a constant.

Letting t_β denote the t-statistic testing $\beta = 0$ in (10) and t_ρ denote the ADF statistic testing $\rho = 1$ in (11) then, under the local to unity specification, t_β and t_ρ have the following joint limiting distribution under the null hypothesis that $\beta = 0$ and $\rho = 1$:

$$(t_\beta, t_\rho) \Rightarrow (\delta \tau_c + (1 - \delta^2)^{1/2} z, \tau_c + c\theta_c) \tag{12}$$

where $\tau_c = (\int J_c^{\mu 2})^{-1/2} \int J_c^\mu dB$, $\theta_c = (\int J_c^{\mu 2})^{1/2}$, B is a Brownian motion, $J_c^\mu(s) = J_c(s) - \int_0^1 J_c(\omega) d\omega$ where the diffusion process J_c is defined by $dJ_c = cJ_c(s)ds + dB(s)$, $J_c(0) = 0$, and z is a standard normal random variable distributed independently of (B, J_c). (See, for example, Stock (1991), especially Appendix A.) Notice that unlike the unit root case, which is nested in (12) for $c = 0$, the limiting distribution now depends on functionals of an Ornstein-Uhlenbeck process governed by c rather than functionals of Brownian motion as in the unit root case.

Asymptotically, only when $\delta = 0$ is t_β normally distributed independently of t_ρ. For nonzero δ, t_β has a nonstandard distribution which depends on both c and δ. However, while δ can be consistently estimated by the sample correlation between $\hat{u}_t$ and $\hat{\nu}_t$, unfortunately, c cannot be. This follows from the fact that although ρ can be consistently estimated, that is,

[14] Local to unity asymptotic distributions provide good approximations to finite sample distributions when the root is close to one. See Chan (1988).

$(\hat{\rho} - \rho) = O_p(1/T)$, if $\hat{\rho} = 1 + \hat{c}/T$ then $\hat{c} - c = O_p(1)$. As a result, an estimator $\hat{c}$ cannot be simply substituted for c when selecting critical values for tests of β.[15]

If we simply ignore the possibility of a unit root in the stochastic explanatory variable when $\delta \neq 0$ then the normal distribution provides a poor approximation to the appropriate nonstandard distribution and will result in substantial size distortions of the predictive regression.[16] Furthermore, a two-step procedure in which we *first* pretest for a unit root in (11) and *then* use standard or nonstandard critical values for inference in (10) depending on whether a unit root is rejected or not, respectively, will also produce large size distortions because of its dependence on the parameter c which cannot be estimated consistently.[17] Intuitively, under the local to unity model the first step test will be unable reject the null hypothesis of a unit root, $c = 0$, with probability one asymptotically even though ρ may in fact be large but less than one, $c < 0$. To the extent that in the first step the stochastic explanatory variable cannot asymptotically be classified correctly, the resultant poor critical values will result in incorrect inference regarding β in the second step.

4.1.1 Bonferroni Intervals

We can, however, construct asymptotically valid tests of $\beta = 0$ in (10) which do *not* depend on c by constructing Bonferroni intervals for β. These are confidence intervals for β in which the dependence on c is eliminated by invoking the Bonferroni inequality. While Bonferroni intervals are conservative, they can be subsequently adjusted so that their nominal size equals a desired level asymptotically.

Let $C_{\varsigma_1}(c) \equiv (c^l_{\varsigma_1}, c^u_{\varsigma_1})$ denote the $100(1 - \varsigma_1)\%$ confidence interval for c that we have previously calculated for a number of stochastic explanatory variables in Table I. A $100(1 - \varsigma_2)\%$ confidence interval for β which depends on c can then be calculated by inverting the corresponding t_β distribution, $C_{\varsigma_2}(\beta \mid c) \equiv (d_{t_\beta,c,\varsigma_2/2}, d_{t_\beta,c,1-\varsigma_2/2})$. Then a Bonferroni confidence interval for β which does <u>not</u> depend on c can be constructed as

$$C^B_\varsigma(\beta) = \bigcup_{c \in C_{\varsigma_1}(c)} C_{\varsigma_2}(\beta \mid c).$$

By the Bonferroni inequality, the interval $C^B_\varsigma(\beta) \equiv (\min_{c^l \leq c \leq c^u} d_{t_\beta,c,\varsigma_2/2}, \max_{c^l \leq c \leq c^u} d_{t_\beta,c,1-\varsigma_2/2})$ has a confidence level of at least $100(1 - \varsigma)\%$ where $\varsigma = \varsigma_1 + \varsigma_2$.

[15] However, see Valkanov (1998) where a consistent estimator of c is derived in the context of testing the rational expectations hypothesis of the term structure.

[16] For example, Elliot and Stock report Monte Carlo evidence that for $\rho = 0.975$, $\delta = -0.9$, and $T = 100$, a one-sided Z-test with nominal level of 5% actually has a rejection rate of 23%.

[17] For example, Cavanagh, Elliott and Stock provide Monte Carlo evidence that for $c = -20$, $\delta = -0.9$, and $T = 500$, the corresponding second step test with nominal level of 5% actually has a rejection rate of 37%.

Table IV tabulates Bonferroni confidence intervals of the slope coefficients obtained by regressing one period *log* real returns of the *VW* index against lagged values of the posited stochastic explanatory variables. The confidence intervals are calculated as

$$\hat{\beta} - d^u(\varsigma_1, \varsigma_2) \times SE(\hat{\beta}) \leq \beta \leq \hat{\beta} - d^l(\varsigma_1, \varsigma_2) \times SE(\hat{\beta}) \tag{13}$$

where we define $d^l(\varsigma_1, \varsigma_2) = \min_{c^l \leq c \leq c^u} d_{t_\beta, c, \varsigma_2/2}$, $d^u(\varsigma_1, \varsigma_2) = \max_{c^l \leq c \leq c^u} d_{t_\beta, c, 1-\varsigma_2/2}$ and $SE(\hat{\beta}) = \{\hat{\sigma}_{\nu\nu}/(\sum_{t=2}^{T} x_{t-1}^2)\}^{\frac{1}{2}}$. We follow Cavanagh, Stock and Elliot and choose ς_1 and ς_2 to adjust the asymptotic size of this Bonferroni test to equal 10%.[18]

The evidence that *log* real dividend yields predict returns in the post-1952 sample period is now seen to be marginal at best. In other words, once we explicitly acknowledge the uncertainty in ρ, there is no reliable evidence that *log* real dividend yields can predict one-period ahead returns.[19] This conclusion is consistent with the results of Table III where we see that the significance of the predictive regression's slope coefficient when using *log* real dividend yields varies with the assumed value of ρ. In this case the cost of not knowing ρ, or equivalently not knowing c, is high because the disturbances to the predictive regression and the autoregression are very correlated, for example, $\hat{\delta} = -0.96$ in the post-1952 sample period.

Similarly, there is no reliable evidence that the *log* book-to-market ratio can forecast one-period ahead returns. All 90% Bonferroni confidence intervals of the slope coefficient obtained by regressing one period *log* real returns of the *VW* index against lagged values of the *log* book-to-market ratio include $\beta = 0$. Again this conclusion is not surprising in light of the results of Table III where the significance of the predictive regression's slope coefficient when using *log* book-to-market ratios varies with the assumed value of ρ, this sensitivity stemming from the fact that the disturbances to the predictive regression and the autoregression are very correlated in this case as well, for example, $\hat{\delta} = -0.81$ for the whole sample period.

In contrast, the term spread is seen to reliably forecast one-period ahead returns in the post-1952 sample period even after explicitly acknowledging the uncertainty in ρ. The fact that when we use the term spread the disturbances to the predictive regression and the autoregression are uncorrelated, $\hat{\delta} = -0.02$ for the post-1952 sample, implies that in this case the cost of not knowing ρ is not high and our inference on β will be robust. This can be

[18] Without this subsequent adjustment, the asymptotic size of the Bonferroni test, $S_B(c, \varsigma_1, \varsigma_2)$, is conservative, $S_B(c, \varsigma_1, \varsigma_2) \leq \varsigma_1 + \varsigma_2$. We construct asymptotically valid confidence intervals having a size of 10% by setting $\varsigma_2 = 10\%$ and numerically determining that value of ς_1 which solves $\sup_c S_B(c, \varsigma_1, 10\%) = 10\%$. This numerical computation is lengthy because we need to compute first stage confidence intervals for each realization of a Bonferroni test statistic. We thank Jim Stock for making his computer procedures for calculating these intervals available to us. Our computations are based on Monte Carlo simulations assuming time series of length $T = 500$ and 20,000 replications over an equally spaced grid of c, $-35 \leq c \leq 5$.

[19] In fact, in unreported calculations we find that the corresponding 95% Bonferroni confidence interval does include $\beta = 0$.

confirmed in Table III where we see that the significance of the predictive regression's slope coefficient when using the term spread does not vary with the assumed value of ρ.

The results of Table IV confirm that the use of near integrated regressors to predict one period ahead returns may have a potentially significant effect on the inferences drawn in these predictive regressions. Our conclusions, however, are particular to the case of one period ahead returns. Recently there has been much more interest in regressions of longer horizon returns onto stochastic explanatory variables. Examples include Fama and French (1988,1989), Keim and Stambaugh (1986) and Hodrick (1992). We now use this local to unity framework to investigate statistical inference when predicting long horizon returns.

4.2 Long Horizon Returns

It is common in predictive regressions to use K $(K > 1)$ period returns

$$y_t(K) = \sum_{i=1}^{K} y_{t+i-1}, \quad t = 1, \ldots, T-K+1$$

rather than one period returns to increase statistical power and improve statistical efficiency. We now assume that this long horizon return, $y_t(K)$, is regressed against the lagged value of the stochastic explanatory variable x_{t-1}:

$$y_t(K) = \alpha(K) + \beta(K)x_{t-1} + u_t(K), \quad t = 1, \ldots, T-K+1. \tag{14}$$

In this section, we investigate within the local-to-unity framework the asymptotic distribution of statistics based on long horizon returns. Like Richardson and Stock (1989) we also use an alternative asymptotic distribution theory for these statistics which recognizes that even though the sample size T is large, the number of nonoverlapping observations may in fact be small resulting in conventional large sample approximations performing poorly in practice. Unlike Richardson and Stock, however, we do not consider the case where long horizon returns are regressed against past returns but rather are regressed against the lagged values of a stochastic explanatory variable whose order of integration, furthermore, is unknown. We follow Phillips (1987a) to derive our asymptotic results and consider a time series x_t generated by the autoregression $x_t = \mu + \rho x_{t-1} + v_t$ with ρ nested in a $1/T$ neighborhood of one, $\rho = exp(c/T)$ where c is a constant.[20]

The maintained null hypothesis is that x_{t-1} cannot be used to predict not only next period's return y_t but also any subsequent one period return $y_t, \ldots, y_{t+K-1}$. As a consequence,

[20]Since $exp(c/T) = 1 + c/T + O(T^{-2})$ our asymptotic results are the same as if we assumed the alternative nesting $\rho = 1 + c/T$.

the resultant K period return $y_t(K)$ will not be predictable using x_{t-1}. Under this null hypothesis $u_t(K) = u_t + u_{t+1} + \ldots + u_{t+K-1}$ because subsequent one period returns are unpredictable using x_{t-1} implying that, apart from a constant, $u_{t+i-1} = y_{t+i-1}$. We complete our specification by assuming that $\epsilon_t = (u_t, v_t)'$ is a martingale difference sequence, $E(\epsilon_t \epsilon_t' \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = \Sigma$ with typical element σ_{uv} and that u_t and v_t are only correlated contemporaneously with $\delta \equiv corr(u_t, v_t)$.

We investigate the asymptotic distributions of the least squares estimator[21] obtained when the long horizon return $y_t(K)$ is regressed against the lagged value of the stochastic explanatory variable x_{t-1}:

$$\hat{\beta}(K) = \left(\sum_{t=1}^{T-K+1} x_{t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T-K+1} y_t(K) x_{t-1} \right)$$

and its corresponding t-statistic. Notice, however, that the residuals $u_t(K)$ of (14) are serially correlated (up to lag $K-1$) under the null hypothesis owing to the overlapping nature of the long horizon returns data. Let $\hat{u}_t(K)$ be the fitted residuals of (14). Then the Newey-West (1987) heteroskedasticity- and autocorrelation-consistent standard error up to lag $K-1$ for $\hat{\beta}(K)$ is:

$$\hat{\sigma}^2_{\hat{\beta}(K)} = \left(\sum_{t=1}^{T-K+1} x_{t-1}^2 \right)^{-2} \left(\sum_{i=-K+1}^{K-1} (1 - |i|/K) \sum_{t=|i|+1}^{T-K+1} \hat{u}_t(K) \hat{u}_{t-|i|}(K) x_{t-1} x_{t-1-|i|} \right)$$

The t-statistic of $\hat{\beta}(K)$ using the Newey-West standard error is then defined by

$$\hat{t}_{\hat{\beta}(K)} = \left(\sum_{i=-K+1}^{K-1} (1 - |i|/K) \sum_{t=|i|+1}^{T-K+1} \hat{u}_t(K) \hat{u}_{t-|i|}(K) x_{t-1} x_{t-1-|i|} \right)^{-\frac{1}{2}} \left(\sum_{t=1}^{T-K+1} y_t(K) x_{t-1} \right).$$

We investigate the behavior of $\hat{\beta}(K)$ and $\hat{t}_{\hat{\beta}(K)}$ under two limiting specifications. The first is the standard large sample approximation where $T \to \infty$ for K fixed so that $K/T \to 0$. The implication of this specification is that as the sample size T increases the amount of independent information also increases. However, to capture the intuition that there is a limit to the amount of independent information in a lengthy time series of overlapping long horizon returns, we also follow Richardson and Stock and let K grow proportionally with T as $T \to \infty$ so that $K/T \to \theta$, or equivalently $K = [T\theta]$ with $[\bullet]$ denoting the greatest lesser integer function, where θ is a fixed constant with $0 < \theta < 1$. Richardson and Stock find that this latter limiting specification provides a better approximation to corresponding finite sample distributions in the case of using long horizon returns to test mean reversion in stock prices.

[21]For expositional purposes we assume, without loss of generality, that $\alpha(K) = 0$ and $\mu = 0$.

The following Lemma provides the results needed to prove our distributional results. In what follows "$\rightarrow$" and "$\Rightarrow$" denote convergence in probability and convergence in distribution, respectively.

Lemma 2

Let W_1 and W_2 be two standard Brownian motions with initial conditions $W_1(0) = W_2(0) = 0$ and correlation $corr(dW_1, dW_2) = \delta$. Define an Ornstein-Uhlenbeck process J_c by $dJ_c(r) = cJ_c(r)dr + dW_1(r)$, where $J_c(0) = 0$. $W_2(r)$ can be rewritten as $W_2(r) = \delta W_1 + (1-\delta^2)^{\frac{1}{2}}\widetilde{W}_2$, where $\widetilde{W}_2$ is a standard Brownian motion distributed independently of W_1.

As $T \rightarrow \infty$,
(a) $T^{-\frac{1}{2}} x_{[Tr]} \Rightarrow \sigma_v J_c(r)$.
If K is fixed, then
(b) $T^{-2}\sum_{t=1}^{T-K+1} x_{t-1}^2 \Rightarrow \sigma_v^2 \int_0^1 J_c(r)^2 dr$;
(c) $T^{-1}\sum_{t=1}^{T-K+1} y_t(K)x_{t-1} \Rightarrow K\sigma_u\sigma_v \left[\delta \int_0^1 J_c(r)dW_1(r) + (1-\delta^2)^{\frac{1}{2}} \int_0^1 J_c(r)d\widetilde{W}_2(r)\right]$.
If $K = [T\theta]$, where $0 < \theta < 1$ is a constant, then
(d) $T^{-2}\sum_{t=1}^{T-K+1} x_{t-1}^2 \Rightarrow \sigma_v^2 \int_0^{1-\theta} J_c(r)^2 dr$;
(e) $T^{-\frac{1}{2}} y_{[Tr]}(K) \Rightarrow \sigma_u[W_2(r+\theta) - W_2(r)]$;
(f) $T^{-2}\sum_{t=1}^{T-K+1} y_t(K)x_{t-1} \Rightarrow \sigma_u\sigma_v \int_0^{1-\theta}[W_2(r+\theta) - W_2(r)]J_c(r)dr$;

Proof: See the Appendix.

Theorem

If K is fixed, then as $T \rightarrow \infty$,
(a) $\hat{\beta}(K) \rightarrow 0$; and
(b) $\hat{t}_{\hat{\beta}(K)} \Rightarrow \delta\left[\int_0^1 J_c(r)^2 dr\right]^{-\frac{1}{2}} \int_0^1 J_c(r)dW_1(r) + (1-\delta^2)^{\frac{1}{2}} z$, where z is a standard normal random variable distributed independently of (W_1, J_c).
If $K = [T\theta]$, where $0 < \theta < 1$ is a constant, then as $T \rightarrow \infty$,
(c) $\hat{\beta}(K) \Rightarrow (\sigma_u/\sigma_v)\left[\int_0^{1-\theta} J_c(r)^2 dr\right]^{-1} \int_0^{1-\theta}[W_2(r+\theta) - W_2(r)]\, J_c(r)dr \equiv \beta_\theta^*$; and
(d) $\hat{t}_{\hat{\beta}(K)} \Rightarrow \left[\int_{-\theta}^{\theta}\int_s^{1-\theta}(1-|s|/\theta)U(r)U(r-|s|)J_c(r)J_c(r-|s|)drds\right]^{-\frac{1}{2}} \int_0^{1-\theta}[W_2(r+\theta)-W_2(r)]J_c(r)dr$, where $U(r) \equiv W_2(r+\theta) - W_2(r) - (\sigma_v/\sigma_u)\beta_\theta^* J_c(r)$.

Proof: See the Appendix.

Under the fixed K limit we see that the ordinary least squares slope estimator $\hat{\beta}(K)$ is consistent while $\hat{t}_{\hat{\beta}(K)}$ has a non-standard distribution depending on δ as well as c. Under the $K/T \rightarrow \theta$ limit, however, $\hat{\beta}(K)$ is no longer consistent. This is to be expected since we have only $1/\theta$ independent observations in the limit and $\hat{\beta}(K)$'s limiting distribution depends on θ as a result. Similarly, $\hat{t}_{\hat{\beta}(K)}$ has a non-standard distribution which now depends on θ.

To further investigate these distributional results, we conduct the following Monte Carlo experiment. For each of $c = -100$, -50, -10, -5, and 0, and for $\sigma_u = 0.1$, $\sigma_v = 0.1$, and $\delta = -0.9$, we generate, under $H_0 : \beta = 0$, time series of length $T = 720$ (as in Richardson and Stock (1989)) of x_t and y_t. We then form K period returns for $K = 1$, 12, 36, 60, and 120, corresponding to $\theta = 1/720$, $1/60$, $1/20$, $1/12$, and $1/6$, respectively, and calculate $\hat{\beta}(K)$ as well as the corresponding Newey-West corrected t-statistic, $\hat{t}_{\hat{\beta}(K)}$. We repeat this experiment 10,000 times and tabulate the resultant sampling distributions of $\hat{\beta}(K)$ and $\hat{t}_{\hat{\beta}(K)}$ in Tables V and VI, respectively.

From Table V we confirm Nelson and Kim's conjecture that the bias in the slope estimator is consistent with Stambaugh's approximation at all horizons *if* one takes the relevant number of observations to be the number of possible nonoverlapping observations rather than the number of actual overlapping observations used in the regression. For example, for $c = -10$ or, equivalently, $\rho = 0.9862$ given $T = 720$, Stambaugh's approximation for $\theta = 1/720$ is 0.0050 while the corresponding median value of $\hat{\beta}(K)$ is 0.0047. For longer horizons we have 0.0596 *vs* 0.0562 ($\theta = 1/60$), 0.167 *vs* 0.179 ($\theta = 1/20$), 0.298 *vs* 0.276 ($\theta = 1/12$), and finally for $\theta = 1/6$ the approximation becomes 0.596 as compared to $\hat{\beta}(K)$'s corresponding median value of 0.528. While Stambaugh's approximation corrects the center of $\hat{\beta}$'s sampling distribution, it does *not* correctly adjust the tails of the distribution which, of course, is the relevant adjustment for purposes of statistical inference. As suggested by the $K/T \rightarrow \theta$ limit theory, $\hat{\beta}(K)$'s sampling distribution depends on θ. In particular, for a given c as θ increases and so the amount of independent information diminishes, the sampling distribution of $\hat{\beta}(K)$ becomes much wider and more skewed right.

These results suggest that it is incorrect to assess the degree of statistical significance in long horizon regressions using conventional asymptotic theory, for example, by checking whether $\hat{\beta}(K)$ lies within 1.645 standard errors of zero (assuming a 10% significance level). This is confirmed in Table VI where we tabulate $\hat{t}_{\hat{\beta}(K)}$'s sampling distribution as a function of both c and θ. Clearly as θ increases for a given c value, more extreme t-statistic values are needed to reject the null hypothesis of no predictability. By way of example, once again for $c = -10$ or, equivalently, $\rho = 0.9862$ given $T = 720$, the 95^{th} percentile of $\hat{t}_{\hat{\beta}(K)}$'s sampling distribution is 2.92 for $\theta = 1/60$ but is 3.56 for $\theta = 1/20$. Consistent with the $K/T \rightarrow \theta$ limit theory, it is also incorrect to take into account the overlapping nature of the long horizon returns data by calculating Newey-West corrected t-statistics and then, to acknowledge uncertainty in ρ, simply use $\hat{t}_{\hat{\beta}(K)}$'s non-standard sampling distribution assuming the data is non-overlapping. From Table VI we see that for each of the assumed c values the sampling distribution of

$\hat{t}_{\hat{\beta}(K)}$ assuming non-overlapping returns data ($\theta = 1/720$) is quite distinct from its sampling distribution for other θ values.

5 Conclusions

This paper has explored statistical inference in predictive regressions. We pay particular attention to the stochastic properties of the posited explanatory variables and demonstrate that the inferences drawn in these regressions can be sensitive to the assumed properties of the explanatory variables, especially their degree of integration. Confidence intervals for the largest autoregressive roots of time series of commonly used explanatory variables are consistent with these roots being very near to unity if not equal to unity. Once we incorporate this uncertainty in the estimation there is little evidence of predictability.

Given how sensitive the inferences drawn in predictive regressions are to the assumed properties of the explanatory variables and the fact that in practice these properties will never be known with certainty underscores the fact that we should be extremely careful in interpreting the results of these predictive regressions. This is especially true in the case of predictive regressions using long-horizon returns where statistical inference depends not only on the explanatory variable's order of integration but also on the length of the horizon itself.

Appendix

Proof of Lemma 1: Write $Y_t = \begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix}$, $X_t' = \begin{pmatrix} X_{1t}' & 0 \\ 0 & X_{2t}' \end{pmatrix}$, and $B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}$. Then the system (8) can be written in a more compact form:

$$Y_t = X_t' B + e_t.$$

The log likelihood function of this system is

$$\mathcal{L}(B, \Omega) = (T/2) \log |\Omega^{-1}| - (1/2) \sum_{t=1}^{T} (Y_t - X_t' B)' \Omega^{-1} (Y_t - X_t' B).$$

Partition Ω conformably with Y as $\Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}$ and define $\Omega_{11\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. It is straightforward to verify that

$$\Omega^{-1} = \begin{pmatrix} I & 0 \\ -\Omega_{22}^{-1}\Omega_{21} & I \end{pmatrix} \begin{pmatrix} \Omega_{11\cdot 2}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix}.$$

Then $\log |\Omega^{-1}| = \log |\Omega_{11\cdot 2}^{-1}| + \log |\Omega_{22}^{-1}|$. The term inside the summation of $\mathcal{L}(B, \Omega)$ can be written as:

$$\begin{aligned} & \begin{pmatrix} Y_{1t} - X_{1t}' B_1 \\ Y_{2t} - X_{2t}' B_2 \end{pmatrix}' \begin{pmatrix} I & 0 \\ -\Omega_{22}^{-1}\Omega_{21} & I \end{pmatrix} \begin{pmatrix} \Omega_{11\cdot 2}^{-1} & 0 \\ 0 & \Omega_{22}^{-1} \end{pmatrix} \begin{pmatrix} I & -\Omega_{12}\Omega_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} Y_{1t} - X_{1t}' B_1 \\ Y_{2t} - X_{2t}' B_2 \end{pmatrix} \\ = \quad & (Y_{1t} - X_{1t}' B_1 - \Omega_{12}\Omega_{22}^{-1}(Y_{2t} - X_{2t}' B_2))' \Omega_{11\cdot 2}^{-1} (Y_{1t} - X_{1t}' B_1 - \Omega_{12}\Omega_{22}^{-1}(Y_{2t} - X_{2t}' B_2)) \\ & + (Y_{2t} - X_{2t}' B_2)' \Omega_{22}^{-1} (Y_{2t} - X_{2t}' B_2). \end{aligned}$$

Then $\mathcal{L}(B, \Omega)$ can be written as the sum of the conditional log likelihood

$$\begin{aligned} (T/2) \log |\Omega_{11\cdot 2}^{-1}| - (1/2) \sum_{t=1}^{T} & (Y_{1t} - X_{1t}' B_1 - \Omega_{12}\Omega_{22}^{-1}(Y_{2t} - X_{2t}' B_2))' \\ & \cdot \Omega_{11\cdot 2}^{-1} (Y_{1t} - X_{1t}' B_1 - \Omega_{12}\Omega_{22}^{-1}(Y_{2t} - X_{2t}' B_2)) \end{aligned}$$

and the marginal log likelihood

$$(T/2) \log |\Omega_{22}^{-1}| - (1/2) \sum_{t=1}^{T} (Y_{2t} - X_{2t}' B_2)' \Omega_{22}^{-1} (Y_{2t} - X_{2t}' B_2).$$

The marginal log likelihood function only depends on B_2 and Ω_{22} which are assumed known. Hence maximizing $\mathcal{L}(B, \Omega)$ is equivalent to maximizing the conditional log likelihood function. It follows that the maximum likelihood estimate of B_1 is equivalent to the corresponding OLS estimate obtained from the posited multiple regression. □

Proof of Lemma 2: (a), (b) are standard and can be found, for example, in Phillips (1987a). (c) is proved by generalizing the method of Phillips (1987a) and using the decomposition $u_t = \delta(\sigma_u/\sigma_v)v_t + (1-\delta^2)^{\frac{1}{2}}\sigma_u \tilde{u}_t$ where $\tilde{u}_t$ is a standard normal variable independent of v_t (see Cavanagh, Elliott, and Stock (1995)).

The proof of (d) is almost the same as that of (b) except that the upper limit of the integral $1-\theta$ reflects the relation $K = [T\theta]$.

(e) is proved by noticing

$$T^{-\frac{1}{2}} y_{[Tr]}(K) = T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]+K} u_t - T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]-1} u_t = T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]+[T\theta]} u_t - T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]-1} u_t,$$

and using the fact that $T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} u_t \Rightarrow W_2(r)$ (see Richardson and Stock (1989)).

(f) is proved easily by applying (a), (e), and the continuous mapping theorem. □

Proof of the Theorem: To prove (a), multiply the numerator and denominator of $\hat{\beta}(K)$ by T^{-2} and apply (b) and (c) of Lemma 2.

To prove (b), multiply the numerator and denominator of $t_{\hat{\beta}(K)}$ by T^{-1} and generalize the method of Stock (1991).[22] Since the Newey-West (1987) procedure requires the number of lags included to correct for autocorrelation and heteroskedasticity increases with the sample size, we construct $\hat{t}_{\hat{\beta}(K)}$ by including at least $max\left(K-1, T^{\frac{1}{5}}\right)$ lags, which equals $T^{\frac{1}{5}}$ as T is sufficiently large.

To prove (c), multiply the numerators and denominators of $\hat{\beta}(K)$ by T^{-2} and apply (e) and (f) of Lemma 2.

To prove (d), multiply the numerators and denominators of $t_{\hat{\beta}(K)}$ by T^{-2} and use the definition of $U(r)$ and the continuous mapping theorem. □

[22] The details are complicated and are not presented for expositional purposes.

References

1. Campbell, John, Andrew Lo, and Craig MacKinlay, 1997, The econometrics of financial markets, Princeton University Press, Princeton, New Jersey.
2. Cavanagh, Christopher, Graham Elliott and James Stock, 1995, "Inference in models with nearly integrated regressors", Econometric Theory 11, 1131-1147.
3. Chan, N., 1988, "On the parameter inference for nearly nonstationary time series", Journal of the American Statistical Association 83: 857-862.
4. Cochrane, John, 1991, "Production-based asset pricing and the link between stock returns and economic fluctuations", Journal of Finance 46, 209-237.
5. Elliott, Graham and James Stock, 1994, "Inference in time series regression when the order of integration of a regressor is unknown", Econometric Theory 10, 672-700.
6. Fama, Eugene, 1984, "Forward and spot exchange rates", Journal of Monetary Economics 14, 319-338.
7. Fama, Eugene, 1991, "Efficient capital markets: II", Journal of Finance 46, 1575-1618.
8. Fama, Eugene and Kenneth French, 1988, "Dividend yields and expected stock returns", Journal of Financial Economics 22, 3-26.
9. Fama, Eugene and Kenneth French, 1989, "Business conditions and expected returns on stocks and bonds", Journal of Financial Economics 25: 23-49.
10. Fuller, Wayne, 1996, Introduction to statistical time series (second edition), John Wiley and Sons, Inc., New York, New York.
11. Hodrick, Robert, 1992, "Dividend yields and expected stock returns: alternative procedures for inference and measurement", Review of Financial Studies 5, 357-386.
12. Keim, Donald and Robert Stambaugh, 1986, "Predicting returns in the stock and bond markets", Journal of Financial Economics 17, 357-390.
13. Kendall, Maurice, 1953, "The analysis of economic time series, part I. Prices", Journal of the Royal Statistical Society 96, 11-25.
14. Kendall, Maurice, 1954, "Note on bias in the estimation of autocorrelation", Biometrika 41, 403-404.

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

16. Marriott, F.H.C. and J. A. Pope, 1954, "Bias in the estimation of autocorrelations", Biometrika 41, 390-402.

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

18. Nelson, Charles and Charles Plosser, 1982, "Trends and random walks in macroeconomic time series: some evidence and implications", Journal of Monetary Economics 10, 129-162.

19. Newey, Whitney and Kenneth West, 1987, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix", Econometrica 55, 703-708.

20. Phillips, Peter C. B., 1987a, "Towards a unified asymptotic theory for autoregression", Biometrika 74, 535-547.

21. Phillips, Peter C. B., 1987b, "Time series regression with a unit root", Econometrica 55, 277-301.

22. Phillips, Peter C. B., 1991, "Optimal inference in cointegrated systems", Econometrica 59, 283-306.

23. Phillips, Peter C. B. and Pierre Perron, 1988, "Testing for a unit root in time series regression", Biometrika 75, 335-346.

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

25. Richardson, Matthew and James Stock, 1989, "Drawing inferences from statistics based on multiyear asset returns", Journal of Financial Economics 25, 323-348.

26. Stambaugh, Robert, 1986, "Bias in regressions with lagged stochastic regressors", unpublished manuscript, University of Chicago, Chicago, Illinois.

27. Stock, James, 1991, "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series", Journal of Monetary Economics, 28, 435-459.

28. Valkanov, Rossen, 1998, "The term structure with highly persistent interest rates", unpublished manuscript, University of California, Los Angeles, California.

Table I

95% Confidence Intervals for the Largest Autoregressive Root of the Stochastic Explanatory Variables

This Table provides 95% confidence intervals for the largest autoregressive root ρ of stochastic explanatory variables typically used in predictive regressions. Dividend Yield is the log real dividend yield and is constructed as in Fama and French (1988). Default Spread is the log of the difference between monthly averaged annualized yields of bonds rated Baa and Aaa by Moody's. Book to Market is the log of Pontiff and Schall's (1998) Dow Jones Industrial Average (DJIA) book-to-market ratio. Term Spread is the difference between annualized yields of Treasury bonds with maturity closest to ten years at month end and three month Treasury bills. The augmented Dickey-Fuller statistic is denoted by ADF and we follow Nelson and Plosser (1982) in determining the maximum lag length k. OLS estimates of ρ based on an $AR(1)$ autoregression, $\hat{\rho}$, are also tabulated as well as the corresponding small sample bias-adjusted estimates, $\hat{\rho}_{adj}$.

Series	Sample Period	k	ADF	95% interval	$\hat{\rho}$	$\hat{\rho}_{adj}$
Dividend Yield	1926:12 - 1994:12	5	-3.30	(0.960,0.996)	0.981	0.986
	1926:12 - 1951:12	1	-2.84	(0.915,1.004)	0.961	0.974
	1952:12 - 1994:12	1	-2.65	(0.956,1.004)	0.978	0.986
Default Spread	1926:12 - 1994:12	2	-2.49	(0.976,1.003)	0.989	0.994
	1926:12 - 1951:12	3	-0.90	(0.984,1.015)	0.994	1.008
	1952:12 - 1994:12	2	-2.50	(0.963,1.004)	0.982	0.989
Book to Market	1926:12 - 1994:08	6	-2.35	(0.977,1.003)	0.989	0.994
	1926:12 - 1951:12	6	-1.60	(0.967,1.013)	0.977	0.991
	1952:12 - 1994:08	6	-1.24	(0.986,1.008)	0.987	0.994
Term Spread	1926:12 - 1994:12	6	-3.57	(0.955,0.992)	0.946	0.950
	1926:12 - 1951:12	6	-3.11	(0.943,0.999)	0.935	0.943
	1952:12 - 1994:12	2	-1.83	(0.957,1.012)	0.971	0.984

Table II

A Monte Carlo Investigation of Nelson and Kim's Randomization Procedure

For each of $\rho = \{0.90, 0.95, 0.99\}$, $\delta = \{-0.5, -0.9\}$, $\sigma_v = \{0.1, 1.0\}$ and $\sigma_u = 1.0$ we generate under $H_0 : \beta = 0$ time series of length $T = 100$ of x_t and y_t according to

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t \\ x_t &= \mu + \rho x_{t-1} + v_t. \end{aligned}$$

We calculate $\hat{\beta}$ and $\hat{\rho}$ by separately applying OLS to each of these equations and retain the resultant residuals $\{\hat{u}_t\}_{t=1}^{100}$ and $\{\hat{v}_t\}_{t=1}^{100}$. We repeat this experiment 5000 times for each parameter combination and summarize $\hat{\beta}$'s sampling distribution by its 10^{th}, 50^{th}, and 90^{th} percentiles. The small sample bias adjusted value β_{adj} obtained from Stambaugh's approximation assuming $\beta = 0$ is also presented for comparison purposes. For each parameter combination we then take the residuals $\{\hat{u}_t\}_{t=1}^{100}$ and $\{\hat{v}_t\}_{t=1}^{100}$ associated with each of the 10^{th}, 50^{th}, and 90^{th} percentiles of the corresponding $\hat{\beta}$ sampling distribution and randomize each grouping of these residuals. Using these randomized residuals we once again generate under $H_0 : \beta = 0$ time series of length $T = 100$ of x_t and y_t and calculate the bias corrected estimate $\check{\beta}$. We repeat this randomization 500 times and summarize $\check{\beta}$'s sampling distribution by its 10^{th}, 50^{th}, and 90^{th} percentiles.

ρ	δ	σ_v	β_{adj}	$\hat{\beta}$ percentiles:	10^{th}	50^{th}	90^{th}	$\check{\beta}$ percentiles	10^{th}	50^{th}	90^{th}
0.90	-0.90	1.0	0.033		-0.027	0.028	0.110	$\hat{\beta}_{10\%}$:	-0.053	0.021	0.103
								$\hat{\beta}_{50\%}$:	-0.035	0.027	0.114
								$\hat{\beta}_{90\%}$:	-0.021	0.032	0.128
	-0.90	0.1	0.333		-0.271	0.275	1.102	$\hat{\beta}_{10\%}$:	-0.529	0.215	1.027
								$\hat{\beta}_{50\%}$:	-0.349	0.267	1.143
								$\hat{\beta}_{90\%}$:	-0.211	0.316	1.281
	-0.50	1.0	0.019		-0.043	0.015	0.088	$\hat{\beta}_{10\%}$:	-0.069	0.009	0.079
								$\hat{\beta}_{50\%}$:	-0.050	0.015	0.095
								$\hat{\beta}_{90\%}$:	-0.035	0.020	0.113
	-0.50	0.1	0.185		-0.425	0.153	0.894	$\hat{\beta}_{10\%}$:	-0.691	0.094	0.788
								$\hat{\beta}_{50\%}$:	-0.498	0.147	0.940
								$\hat{\beta}_{90\%}$:	-0.354	0.200	1.126
0.95	-0.90	1.0	0.035		-0.013	0.032	0.103	$\hat{\beta}_{10\%}$:	-0.013	0.023	0.097
								$\hat{\beta}_{50\%}$:	-0.024	0.028	0.108
								$\hat{\beta}_{90\%}$:	-0.042	0.034	0.122
	-0.90	0.1	0.347		-0.130	0.312	1.002	$\hat{\beta}_{10\%}$:	-0.104	0.230	0.970
								$\hat{\beta}_{50\%}$:	-0.237	0.283	1.081
								$\hat{\beta}_{90\%}$:	-0.419	0.337	1.216
	-0.50	1.0	0.019		-0.030	0.016	0.078	$\hat{\beta}_{10\%}$:	-0.025	0.010	0.072
								$\hat{\beta}_{50\%}$:	-0.039	0.016	0.087
								$\hat{\beta}_{90\%}$:	-0.057	0.021	0.104
	-0.50	0.1	0.193		-0.294	0.163	0.813	$\hat{\beta}_{10\%}$:	-0.248	0.102	0.717
								$\hat{\beta}_{50\%}$:	-0.389	0.155	0.866
								$\hat{\beta}_{90\%}$:	-0.581	0.210	1.040

Table II (continued)

ρ	δ	σ_v	β_{adj}	$\hat{\beta}$ percentiles:	10^{th}	50^{th}	90^{th}	$\breve{\beta}$ percentiles	10^{th}	50^{th}	90^{th}
0.99	-0.90	1.0	0.036		0.000	0.036	0.100	$\hat{\beta}_{10\%}$:	-0.004	0.015	0.064
								$\hat{\beta}_{50\%}$:	-0.015	0.026	0.096
								$\hat{\beta}_{90\%}$:	-0.032	0.035	0.113
	-0.90	0.1	0.357		0.006	0.367	1.021	$\hat{\beta}_{10\%}$:	-0.004	0.146	0.642
								$\hat{\beta}_{50\%}$:	-0.151	0.262	0.965
								$\hat{\beta}_{90\%}$:	-0.330	0.353	1.132
	-0.50	1.0	0.020		-0.016	0.019	0.072	$\hat{\beta}_{10\%}$:	-0.016	0.007	0.049
								$\hat{\beta}_{50\%}$:	-0.028	0.014	0.074
								$\hat{\beta}_{90\%}$:	-0.046	0.021	0.093
	-0.50	0.1	0.199		-0.165	0.191	0.708	$\hat{\beta}_{10\%}$:	-0.158	0.072	0.500
								$\hat{\beta}_{50\%}$:	-0.280	0.141	0.734
								$\hat{\beta}_{90\%}$:	-0.453	0.207	0.922

Table III

Predictive Regression Results if there is No Uncertainty About the Stochastic Explanatory Variable's Largest Autoregressive Root

This Table provides t-statistics to assess the statistical significance of the slope coefficient in regressing one-month ahead log real returns of the CRSP VW index against a stochastic explanatory variable assuming the explanatory variable's largest autoregressive root ρ is known with certainty. The explanatory variables used are Dividend Yield, Default Spread, Book to Market, and Term Spread and are described in Table I. The columns labeled $\hat{\rho}$ and $\hat{\rho}_{adj}$ give these estimation results assuming that the explanatory variable's largest autoregressive root is given by its historical estimate and its bias-adjusted historical estimate, respectively. The column labeled $\rho = 1$ assumes that the explanatory variable has a unit root. For comparison purposes, the column labeled Predictive Regression presents the t-statistic from a regression against a stochastic explanatory variable *without* incorporating information about the explanatory variable's stochastic properties and corresponds to the standard implementation of a predictive regression. The sample correlation coefficient between this latter regression's estimated disturbances and the estimated disturbances of an $AR(1)$ autoregression in the explanatory variables is denoted by $\hat{\delta}$. Throughout t-statistics are based on Newey-West standard errors.

Explanatory Variable	Sample Period	Predictive Regression	$\hat{\delta}$	$\rho = \hat{\rho}$	$\rho = \hat{\rho}_{adj}$	$\rho = 1$
Dividend Yield	1926:12 - 1994:12	t= 1.338	-0.955	t= 2.691	t= 1.701	t=-1.053
	1926:12 - 1951:12	t= 0.751	-0.952	t= 1.773	t= 0.254	t=-2.607
	1952:12 - 1994:12	t= 2.993	-0.962	t=11.752	t= 8.251	t= 1.871
Default Spread	1926:12 - 1994:12	t= 0.337	-0.266	t= 0.372	t= 0.271	t= 0.131
	1926:12 - 1951:12	t= 0.044	-0.391	t= 0.059	t= -0.602	t=-0.229
	1952:12 - 1994:12	t= 1.247	0.027	t= 1.234	t= 1.257	t= 1.288
Book to Market	1926:12 - 1994:08	t= 1.645	-0.811	t= 3.615	t= 2.838	t= 0.344
	1926:12 - 1951:12	t= 1.386	-0.888	t= 5.297	t= 2.463	t= 0.693
	1952:12 - 1994:08	t= 0.902	-0.699	t= 1.138	t= 0.252	t=-0.487
Term Spread	1926:12 - 1994:12	t= 1.317	-0.052	t= 1.350	t= 1.324	t= 1.032
	1926:12 - 1951:12	t=-0.422	-0.125	t=-0.449	t=-0.567	t=-0.698
	1952:12 - 1994:12	t= 2.399	-0.020	t= 2.420	t= 2.407	t= 2.242

Table IV

90% Bonferroni Confidence Intervals

This Table provides 90% Bonferroni confidence intervals to assess the statistical significance of the slope coefficient in regressing one-month ahead log real returns of the CRSP VW index against a stochastic explanatory variable assuming the explanatory variable's largest autoregressive root ρ is characterized by a local-to-unity specification $\rho = 1 + c/T$ where c is a nuisance parameter and T denotes the sample size. The explanatory variables used are Dividend Yield, Default Spread, Book to Market, and Term Spread and are described in Table I. The intervals dependence on c is eliminated by invoking the Bonferroni inequality and are subsequently adjusted so that their nominal size equals 90% asymptotically.

Explanatory Variable	Sample Period	90% Confidence Interval for $\hat{\beta}$
Dividend Yield	1926:12 - 1994:12	(-0.002,0.021)
	1926:12 - 1951:12	(-0.023,0.033)
	1952:12 - 1994:12	(0.002,0.035)
Default Spread	1926:12 - 1994:12	(-0.005,0.009)
	1926:12 - 1951:12	(-0.015,0.009)
	1952:12 - 1994:12	(0.000,0.015)
Book to Market	1926:12 - 1994:08	(-0.005,0.009)
	1926:12 - 1951:12	(-0.010,0.031)
	1952:12 - 1994:08	(-0.009,0.012)
Term Spread	1926:12 - 1994:12	(-0.001,0.005)
	1926:12 - 1951:12	(-0.010,0.005)
	1952:12 - 1994:12	(0.001,0.006)

Table V

Sampling Properties of the Estimated Long Horizon Slope Coefficient $\hat{\beta}(K)$ under the Local-to-Unity Specification

For $\sigma_u = 0.1$, $\sigma_v = 0.1$ and $\delta = -0.9$ we generate under $H_0 : \beta = 0$ time series of length $T = 720$ of x_t and y_t according to

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t \\ x_t &= \mu + \rho x_{t-1} + v_t \end{aligned}$$

assuming that the explanatory variable's largest autoregressive root ρ is characterized by a local-to-unity specification $\rho = 1 + c/T$ for $c = \{-100, -50, -10, -5, 0\}$. We then form K period returns, $y_t(K)$, $K = 1, 12, 36, 60$, and 120, corresponding to $\theta \equiv K/T = 1/720, 1/60, 1/20, 1/12$, and $1/6$, respectively, and the long horizon regression coefficient $\hat{\beta}(K)$ is obtained by regressing $y_t(K)$ against x_{t-1}. We repeat this experiment 10,000 times and summarize $\hat{\beta}(K)$'s sampling distribution by its 2.5%, 5%, 10%, 50%, 90%, 95%, and 97.5% percentiles.

		2.5%	5%	10%	50%	90%	95%	97.5%
c=-100	θ=1/720	-0.0312	-0.0262	-0.0198	0.0034	0.0298	0.0379	0.0454
	θ=1/60	-0.2958	-0.2449	-0.1823	0.0385	0.2650	0.3272	0.3784
	θ=1/20	-0.6426	-0.5253	-0.3623	0.1227	0.5477	0.6523	0.7440
	θ=1/12	-0.8573	-0.6703	-0.4594	0.2072	0.7172	0.8456	0.9632
	θ=1/6	-1.2001	-0.8947	-0.5374	0.3987	1.0344	1.2118	1.3503
c=-50	θ=1/720	-0.0204	-0.0169	-0.0129	0.0036	0.0239	0.0304	0.0367
	θ=1/60	-0.2146	-0.1811	-0.1337	0.0395	0.2352	0.2949	0.3451
	θ=1/20	-0.5362	-0.4342	-0.3155	0.1294	0.5449	0.6469	0.7452
	θ=1/12	-0.7732	-0.6049	-0.4229	0.2210	0.7348	0.8740	0.9868
	θ=1/6	-1.1230	-0.8356	-0.5172	0.4217	1.0824	1.2507	1.3911
c=-10	θ=1/720	-0.0067	-0.0054	-0.0038	0.0043	0.0166	0.0211	0.0247
	θ=1/60	-0.0790	-0.0636	-0.0426	0.0506	0.1868	0.2334	0.2762
	θ=1/20	-0.2308	-0.1802	-0.1169	0.1537	0.4887	0.5963	0.6849
	θ=1/12	-0.3710	-0.2789	-0.1737	0.2526	0.7213	0.8490	0.9574
	θ=1/6	-0.6196	-0.4406	-0.2273	0.4847	1.1085	1.2739	1.4023
c=-5	θ=1/720	-0.0041	-0.0031	-0.0018	0.0047	0.0155	0.0196	0.0235
	θ=1/60	-0.0476	-0.0365	-0.0204	0.0562	0.1775	0.2219	0.2607
	θ=1/20	-0.1439	-0.1026	-0.0529	0.1673	0.4827	0.5814	0.6593
	θ=1/12	-0.2331	-0.1668	-0.0791	0.2764	0.7148	0.8396	0.9481
	θ=1/6	-0.4100	-0.2568	-0.0854	0.5282	1.1248	1.2816	1.4113
c=0	θ=1/720	-0.0011	-0.0002	0.0007	0.0053	0.0143	0.0180	0.0216
	θ=1/60	-0.0129	-0.0029	0.0086	0.0633	0.1656	0.2086	0.2422
	θ=1/20	-0.0413	-0.0109	0.0244	0.1869	0.4549	0.5537	0.6357
	θ=1/12	-0.0748	-0.0193	0.0398	0.3069	0.6974	0.8163	0.9243
	θ=1/6	-0.1750	-0.0557	0.0671	0.5736	1.1215	1.2757	1.4014

Table VI

Sampling Properties of the Newey-West Corrected t-statistic of the Estimated Long Horizon Slope Coefficient $\hat{\beta}(K)$ under the Local-to-Unity Specification

For $\sigma_u = 0.1$, $\sigma_v = 0.1$ and $\delta = -0.9$ we generate under $H_0 : \beta = 0$ time series of length $T = 720$ of x_t and y_t according to

$$\begin{aligned} y_t &= \alpha + \beta x_{t-1} + u_t \\ x_t &= \mu + \rho x_{t-1} + v_t \end{aligned}$$

assuming that the explanatory variable's largest autoregressive root ρ is characterized by a local-to-unity specification $\rho = 1 + c/T$ for $c = \{-100, -50, -10, -5, 0\}$. We then form K period returns, $y_t(K)$, $K = 1, 12, 36, 60$, and 120, corresponding to $\theta \equiv K/T = 1/720, 1/60, 1/20, 1/12$, and $1/6$, respectively, and the long horizon regression coefficient $\hat{\beta}(K)$ is obtained by regressing $y_t(K)$ against x_{t-1} and calculating the Newey-West corrected t-statistic, $\hat{t}_{\hat{\beta}(K)}$. We repeat this experiment 10,000 times and summarize $\hat{t}$'s sampling distribution by its 2.5%, 5%, 10%, 50%, 90%, 95%, and 97.5% percentiles.

		2.5%	5%	10%	50%	90%	95%	97.5%
c=-100	θ=1/720	-1.8115	-1.5058	-1.1102	0.1732	1.4542	1.8137	2.1454
	θ=1/60	-1.9999	-1.6342	-1.2269	0.2581	1.8955	2.4329	2.8796
	θ=1/20	-1.7942	-1.4429	-1.0272	0.4069	2.2674	2.8753	3.4421
	θ=1/12	-1.7028	-1.3396	-0.9696	0.5585	2.5219	3.2792	3.8805
	θ=1/6	-1.6513	-1.3094	-0.8538	0.9172	3.1851	3.9514	4.6897
c=-50	θ=1/720	-1.7152	-1.3674	-1.0187	0.2579	1.5233	1.8966	2.2069
	θ=1/60	-2.0229	-1.6370	-1.2045	0.3303	2.0072	2.5409	2.9931
	θ=1/20	-1.9012	-1.5302	-1.0775	0.4899	2.3766	3.0047	3.6479
	θ=1/12	-1.8606	-1.4511	-1.0216	0.6328	2.7104	3.4898	4.2188
	θ=1/6	-1.8022	-1.3739	-0.9046	1.0017	3.4462	4.3084	5.1599
c=-10	θ=1/720	-1.2835	-0.9847	-0.6475	0.5833	1.8281	2.1627	2.4399
	θ=1/60	-1.6158	-1.2438	-0.7907	0.7680	2.4037	2.9174	3.3551
	θ=1/20	-1.7034	-1.2934	-0.7844	0.9049	2.8513	3.5568	4.2339
	θ=1/12	-1.7803	-1.2774	-0.7609	1.0491	3.3645	4.2269	5.1197
	θ=1/6	-1.8843	-1.2723	-0.6501	1.4971	4.6431	5.7978	7.1590
c=-5	θ=1/720	-1.0217	-0.7435	-0.4095	0.7966	1.9843	2.3297	2.6042
	θ=1/60	-1.2919	-0.8970	-0.4978	1.0415	2.6480	3.1251	3.5450
	θ=1/20	-1.3414	-0.9324	-0.4488	1.1976	3.1681	3.8801	4.5358
	θ=1/12	-1.4004	-0.9631	-0.4271	1.3553	3.7527	4.7059	5.4760
	θ=1/6	-1.5687	-0.9316	-0.2801	1.8928	5.2522	6.5734	7.9746
c=0	θ=1/720	-0.4224	-0.0804	0.2599	1.3946	2.4837	2.7902	3.0609
	θ=1/60	-0.5011	-0.1142	0.3151	1.7845	3.3132	3.7864	4.2491
	θ=1/20	-0.5656	-0.1517	0.3229	2.0008	3.9933	4.6933	5.3488
	θ=1/12	-0.6938	-0.1852	0.3446	2.2230	4.7266	5.7568	6.7456
	θ=1/6	-0.9931	-0.3247	0.3475	2.8997	6.8444	8.5979	10.7191