

A NONPARAMETRIC TEST OF THE PREDICTIVE REGRESSION MODEL

TED JUHL

ABSTRACT. Predictive regression models are often used in finance to model stock returns as a function of some observable variables. This paper considers testing the significance of a regressor with a unit root in a predictive regression model. The procedures discussed in this paper are nonparametric, so one can test the significance of a regressor without specifying a functional form. We show that the test has a normal distribution, just as in the case where the regressor is stationary. A Monte Carlo experiment explores the performance of the test for various levels of persistence of the regressors and for various linear and nonlinear alternatives. The test has superior performance against certain nonlinear alternatives. An application of the test applied to stock returns shows how the test can improve inference about predictability.

1. INTRODUCTION

The predictive regression model has received much attention in recent years. In particular, the model where the regressors are persistent has been covered in the work of Campbell and Yogo (2006), Cavenagh, Elliott, and Stock (1995), Jansson and Moreira (2006), Lewellen (2004), and Stambaugh (1999), among others. The statistical estimation and testing of such models requires a treatment beyond the traditional normal approximations for the regression parameters. The persistence of the regressors directly affects the distribution of the standard t-ratio of the estimated regression parameters. The resulting distribution of the regression parameter is related to the Dickey-Fuller distribution but with a nuisance parameter from the correlation of innovations to

Version: January 23, 2007. I thank Michael Jansson for providing the code for his procedures. Ted Juhl, Department of Economics, University of Kansas, 415 Snow, Lawrence, KS 66045; Tel (785)-864-2849; (juhl@ku.edu).

the unit root series in the future. The distribution is non-normal so that standard asymptotic theory does not apply. Moreover, as the amount of feedback increases, the distribution shifts to the right so that a type of “spurious” regression further distorts standard inference. These facts indicate the need for knowledge of the persistence of the regressor, or the need to develop new procedures that are “robust” to the unknown level of persistence in the regressor.

Given the above problems, Campbell and Yogo (2006), Jansson and Moreira (2006), and Lewellen (2004) each propose solutions to inference when the level of persistence of the regressor is unknown and possibly contains a unit root. The approaches are related in that they attempt to find an optimal test in some class of statistics that are invariant to location and scale transformations of the data. The test we propose in this paper shares the invariance properties of the above statistics. However, our test is designed to have power against linear and nonlinear alternatives. To accomplish this, we use nonparametric kernel estimation of the regression function. The estimated function is then tested to see if it is significantly different from zero. In this way, we have power against a much wider class of alternative hypotheses. Of course, with such generality comes a price in that power is not as high if there truly exists a linear relationship. We explore these costs via a Monte Carlo experiment. We find that the nonparametric test does outperform tests based on the linear for certain nonlinear alternatives.

The remainder of the paper is organized as follows. In section 2, we describe the predictive regression model. Section 3 provides an outline of the nonparametric test. The asymptotic results are derived in Section 4 and explored in a Monte Carlo experiment detailed in Section 5. An empirical illustration using stock returns is treated in Section 6 and Section 7 concludes. Notation is standard with weak convergence denoted by \xrightarrow{d} and convergence in probability by \xrightarrow{p} . Integrals with respect to Lebesgue measure such as $\int_0^1 ds$ are usually

written as $\int ds$ when there is no ambiguity over limits. All limits in the paper are taken as the sample size $T \rightarrow \infty$.

2. PREDICTIVE REGRESSION MODEL

Let the variable y_t denote the dependent variable in period t and let x_{t-1} denote a predictor variable observed at time $t - 1$. We are interested to see if there is any relationship between y_t and x_{t-1} . Without loss of generality, we suppose that y_t has mean zero so that it can be interpreted as deviations from a time average over the sample period considered. We consider a model of the form

$$\begin{aligned} y_t &= g(x_{t-1}) + \epsilon_t \\ x_t &= x_{t-1} + e_t. \end{aligned}$$

Note that in this case, the variable x_t has a unit root. This is an extreme form of dependence in the predictor variable but one that serves as an approximation for the behavior of many financial ratios used in this type of analysis. Several studies have versions of this model where $g(x_{t-1})$ is considered a linear function. and the findings suggest that the distribution of OLS in the linear case is non-standard and depends on the correlation between ϵ_t and e_t . Moreover, if the series x_t does not have a unit root but instead has a “near unit root”, so that the autoregressive parameter gets close to one as the sample size increases, the model has another nuisance parameter, one that can’t be consistently estimated. In light of these problems, Jansson and Moreira (2006), Campbell and Yogo (2006), Lewellen (2004), and others have proposed tests that are not based on the usual t-ratio of the OLS squares estimator of the slope coefficient for x_{t-1} . For Jansson and Moreira (2006), the distribution of their conditional test is also nonstandard and requires a numerical integration routine to calculate p-values for the test.

In this paper, we propose testing the predictive regression model using the nonparametric tests described in the next section.

3. NONPARAMETRIC TESTS

In this section, we summarize the type of nonparametric tests that we employ in this paper. This type of test has been used when the data is assumed to be stationary as in Fan and Li (1996, 1999) and Zheng (1996) among others. The idea is to check if two series are related without specifying *how* they are related. That is, we allow for the series to have an unknown, perhaps nonlinear, relationship. To be specific, suppose that y_t and z_t are stationary series and that

$$(3.1) \quad y_t = g(z_t) + \epsilon_t$$

where $E(\epsilon_t|z_t) = 0$ which implies that z_t and ϵ_t are uncorrelated and that $E(\epsilon_t) = 0$. To simplify things, suppose that y_t has mean zero. The model then implies that $E(y_t|z_t) = g(z_t)$. The hypothesis that we wish to test is

$$H_0 : g(z_t) = 0$$

versus the alternative

$$H_1 : g(z_t) \neq 0.$$

Let $p(z_t)$ be the density function of z_t . Then using iterated expectations, we have

$$(3.2) \quad E(y_t E(y_t|z_t) p(z_t)) = E(E(y_t|z_t)^2 p(z_t)) \geq 0$$

with equality to zero in (3.2) if and only if $E(y_t|z_t) = g(z_t) = 0$. The nonparametric tests estimate $E(y_t|z_t)p(z_t)$ by kernel regression of y_t on z_t multiplied by the density estimate of $p(z_t)$. That is

$$\hat{E}(y_t|z_t) = \frac{1}{\hat{p}(z_t)} \frac{1}{Th} \sum_{s \neq t}^T K\left(\frac{z_t - z_s}{h}\right) y_s$$

and

$$\hat{E}(y_t|z_t)\hat{p}(z_t) = \frac{1}{Th} \sum_{s \neq t}^T K\left(\frac{z_t - z_s}{h}\right) y_s$$

where $K(\cdot)$ is a kernel function and h is a bandwidth parameter.¹ The final estimate of $E(y_t E(y_t|z_t) p(z_t))$ is the statistic

$$U_T = \frac{1}{T^2 h} \sum_{t=1}^T \sum_{s \neq t}^T K\left(\frac{z_t - z_s}{h}\right) y_t y_s.$$

When the statistic U_T is properly scaled, it is normally distributed under the null hypothesis that $g(z_t) = 0$. Due to its nonparametric construction, the test has power against a wide variety of alternatives, both linear and nonlinear.

4. DISTRIBUTION THEORY

We list the assumptions in our model.

Assumption 1. *The vector*

$$\begin{pmatrix} \epsilon_t \\ e_t \end{pmatrix}$$

is i.i.d. with an elliptical distribution with covariance matrix

$$\begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon e} \\ \sigma_{\epsilon e} & \sigma_e^2 \end{pmatrix}$$

and $E(|e_t|^{\delta_1}) < \infty$ for some $\delta_1 > 2$.

Assumption 2. *The kernel function $K(\cdot)$ is a bounded, symmetric, and non-negative density with an integrable characteristic function and satisfies the*

¹For an introduction to kernel regression and nonparametric regression see Härdle (1989).

following conditions for some $\delta_2 > 2$:

$$\begin{aligned}\int_{-\infty}^{\infty} K(s)ds &= 1 \\ \int_{-\infty}^{\infty} K(s)^2 ds &< \infty \\ \int_{-\infty}^{\infty} s^{2\delta_2} K(s)ds &< \infty.\end{aligned}$$

Assumption 3. *The bandwidth parameter $h \rightarrow 0$ but $Th^2 \rightarrow \infty$.*

The assumption of an elliptical distribution for ϵ_t implies that the distribution is symmetric. Symmetry is a common assumption in models with nonparametric estimation and unit root behavior of the series. For example, symmetry is assumed in Shin and So (1999). Moreover, Jansson (2006) shows that symmetry is a necessary condition for adaptive estimation in unit root models where efficiency gains are possible due to the estimation of unknown error densities. In addition, capital asset pricing models (CAPM) in finance can be obtained using elliptical assumptions about the data. In particular, Owen and Rabinovich (1983) show that elliptical distributions are sufficient to generate CAPM models and Berk (1997) shows that elliptical distributions are also necessary for CAPM. Hodgson, Linton, Vorkink (2002) use a semi-parametric model that imposes elliptical symmetry to test CAPM. The class of elliptical distributions contains the normal distribution, but it also includes the t-distribution and mixed normal distributions so that fat tailed distributions are allowed.

The kernel and bandwidth assumptions are common and are identical to those used in Park and Phillips (1998). The bandwidth assumption guarantees that the data is viewed through a shrinking window as we have more data and that this window does not shrink too fast.

Consider the following variable constructed from a partial sum of the e_t variables:

$$u = e_1 + e_2 + \cdots + e_t.$$

It is easy to show that u will have an elliptical distribution with variance $\sigma_u^2 = \sigma_\epsilon^2 t$. Elliptical distributions have density of the form

$$f_u(u) = \frac{1}{\sigma_u} m_t \left(\frac{u^2}{\sigma_u^2} \right)$$

Notice that the argument of m_t is always squared which generates the symmetry. The subscript t is used to emphasize the dependence of the distribution to the number of terms in the partial sum. We make the following assumption about the function m_t .

Assumption 4. *Suppose that m_t is bounded and has a bounded derivative for each t .*

As a simple example, for the normal distribution, the function is $m_t(w) \sim e^{-w/2t}$ which is clearly bounded and has bounded first derivative.

Theorem 4.1. *Given Assumptions 1 to 4, under the null hypothesis that $g(x_{t-1}) = 0$, the statistic*

$$\frac{T^{5/4} \sqrt{h} U_T}{\sqrt{\hat{\Sigma}_T}} \xrightarrow{d} N(0, 1)$$

where

$$U_T = \frac{1}{T^2 h} \sum_{t=2}^T \sum_{s=1}^{t-1} K \left(\frac{x_{t-1} - x_s}{h} \right) y_t y_s$$

and

$$\hat{\Sigma}_T = \frac{1}{T^{3/2} h} \sum_{t=2}^T \sum_{s=1}^{t-1} K \left(\frac{x_{t-1} - x_s}{h} \right)^2 y_t^2 y_s^2.$$

The theorem states that the standardized statistic has a normal distribution under the null hypothesis that x_{t-1} is unrelated to y_t . This fact is unusual in that standard OLS tests in predictive regression models have a nonstandard limiting distribution if x_{t-1} has a unit root or near unit root. Moreover, when using OLS, the distribution depends on nuisance parameters such as the correlation between ϵ_t and e_t . In our case, the limiting distribution does not change if the series x_{t-1} is stationary or has a unit root.

One difference between our result and the results of Zheng (1996) or Fan and Li (1999) is that the statistic U_T itself does not have a normal distribution although the *standardized* version used in our theorem does have a normal distribution. In the proof of the theorem in the appendix, it is shown that this version of U_T has a mixed normal distribution with a random variance. However, upon scaling by $\hat{\Sigma}_T$, the standardized version of U_T obtains a standard normal distribution.

Using the results of Bandi (2004), it is easy to see that the test is consistent. That is, if the null is not true, the nonparametric estimator will converge to the true nonparametric function, which is nonzero, and then our estimate of $E(E(y_t|x_{t-1})^2 p(x_{t-1}))$ will necessarily be greater than zero and will be multiplied by $T^{5/4}\sqrt{h}$.

From a practical point of view, the test is a good complement to the existing tests based on the linear model in that it may detect dependence in x_{t-1} and y_t that is nonlinear, and the tests are not affected by assumptions about the dependence in x_{t-1} .

One important consideration of nonparametric regression is the choice of bandwidth parameters. The assumptions used in Theorem 4.1 only require that h goes to zero at some prescribed rate. No optimal bandwidth is proposed in this paper but we suggest using $h = d\hat{\sigma}_x T^{-1/5}$ where d is a scalar and $\hat{\sigma}_x$ is the sample standard deviation of x_{t-1} . In our simulations, we try various values of d to gauge the sensitivity of the tests to the bandwidth parameter. One motivation for using a multiple of $\hat{\sigma}_x$ is that the statistic will become scale invariant. The test is always location invariant but using a bandwidth that is a multiple of $\hat{\sigma}_x$ ensures scale invariance.

5. MONTE CARLO EVIDENCE

We compare the nonparametric test with the two leading tests developed by Campbell and Yogo (2006) and Jansson and Moriera (2006). The experiment

uses the following model:

$$y_t = g(x_{t-1}) + \epsilon_t$$

$$x_t = \rho_T x_{t-1} + e_t.$$

We set the parameter $\sigma_{\epsilon e} = -0.75$ and -0.95 , values that are consistent with the empirical regularities found in financial data. For example, if x_{t-1} represents a dividend-price ratio and y_t represents a stock return, then a positive shock to returns in the form of a large value of ϵ_t would then cause a concurrent decrease in the dividend price ratio, so that the correlation between e_t and ϵ_t should be negative. Lewellen (2004) estimates values of $\sigma_{\epsilon e}$ ranging from -0.88 to -0.96 for various financial ratios used for x_{t-1} .

We consider several possible levels of persistence in x_t through the parameter ρ . In particular, we parameterize ρ as

$$\rho_T = 1 - \frac{c}{T}$$

so that ρ_T is “local to unity”. The values of c used in the simulation are 0,5,10,15, and 20. In this first experiment, we examine the size properties of the test. That is, we generate the data so that $g(x_{t-1}) = 0$ and we tabulate the percentage of rejections in 1000 replications. The sample size is given as $T = 100$. There are three different bandwidth parameters that are examined for the U_T statistic, $d = 1, 2$, and 4 in the bandwidth $h = d\hat{\sigma}_x T^{-1/5}$ and we denote the statistic as U1, U2, and U4. The statistic of Jansson and Moriera (2006) is a uniformly most powerful conditionally unbiased test in the linear model, so it is denoted UMPCU. The Campbell and Yogo test uses a Bonferroni bound and we denote this test as CYB. For comparison purposes, we include an OLS t-test that the linear regression coefficient on x_{t-1} is zero and denote this as TSTAT. This test is incorrect in the case when x_t has a unit root or root local to unity, but is included to illustrate the potential problems of ignoring the extreme dependence in x_{t-1} . The size of the tests appear in Table 1.

Table 1.
Size

	c	UMPCU	U1	U2	U4	TSTAT	CYB
$\sigma_{\epsilon e} = -0.95$	$c=0$	0.049	0.034	0.020	0.014	0.394	0.046
	$c=5$	0.040	0.038	0.016	0.006	0.162	0.057
	$c=10$	0.060	0.025	0.009	0.002	0.120	0.049
	$c=15$	0.049	0.036	0.018	0.002	0.079	0.038
	$c=20$	0.055	0.031	0.019	0.002	0.088	0.041
$\sigma_{\epsilon e} = -0.75$	$c=0$	0.050	0.027	0.014	0.006	0.286	0.047
	$c=5$	0.048	0.030	0.014	0.003	0.137	0.035
	$c=10$	0.051	0.030	0.020	0.009	0.088	0.031
	$c=15$	0.064	0.038	0.019	0.007	0.096	0.036
	$c=20$	0.061	0.032	0.018	0.008	0.088	0.033

From Table 1, we see that the UMPCU and CYB tests has size close to the nominal size in all cases. The U statistic is conservative and gets more conservative for a larger bandwidth. As discussed earlier, the t-statistic is not normally distributed and we included this statistic to show the costs of ignoring the unit root behavior of x_{t-1} . At a nominal size of 5%, we reject the null hypothesis of no predictability approximately 40% of the time. When the parameter c increases (near unit root), the normal approximation becomes less damaging. However, even when $c = 5$ with an autoregressive parameter of $\rho_T = 0.95$, we reject the null 16% of the time.

One of the assumptions is that the distribution of ϵ and e is jointly elliptical symmetric. In order to determine the sensitivity of the U statistic to this assumption, we generate the same data using a centered χ_1^2 distribution which is clearly skewed. The experiment is repeated and the results are shown in Table 2.

Table 2.
Size: Chi-square Errors

	c	UMPCU	U1	U2	U4	TSTAT	CYB
$\sigma_{ee} = -0.95$	c=0	0.054	0.022	0.019	0.010	0.388	0.049
	c=5	0.059	0.013	0.008	0.002	0.162	0.043
	c=10	0.055	0.017	0.011	0.003	0.110	0.035
	c=15	0.052	0.016	0.010	0.000	0.089	0.050
	c=20	0.067	0.025	0.013	0.001	0.088	0.043
$\sigma_{ee} = -0.75$	c=0	0.058	0.038	0.020	0.014	0.274	0.034
	c=5	0.045	0.021	0.012	0.003	0.127	0.025
	c=10	0.074	0.028	0.019	0.004	0.099	0.036
	c=15	0.057	0.031	0.016	0.005	0.082	0.037
	c=20	0.050	0.026	0.013	0.002	0.069	0.037

The results in Table 2 suggest that the normal distribution is still a reasonable approximation even if the errors are not from an elliptically symmetric distribution.

For power comparisons, we now consider two alternative functions. In the first case, the alternative is linear so that

$$g(x_{t-1}) = \beta x_{t-1}$$

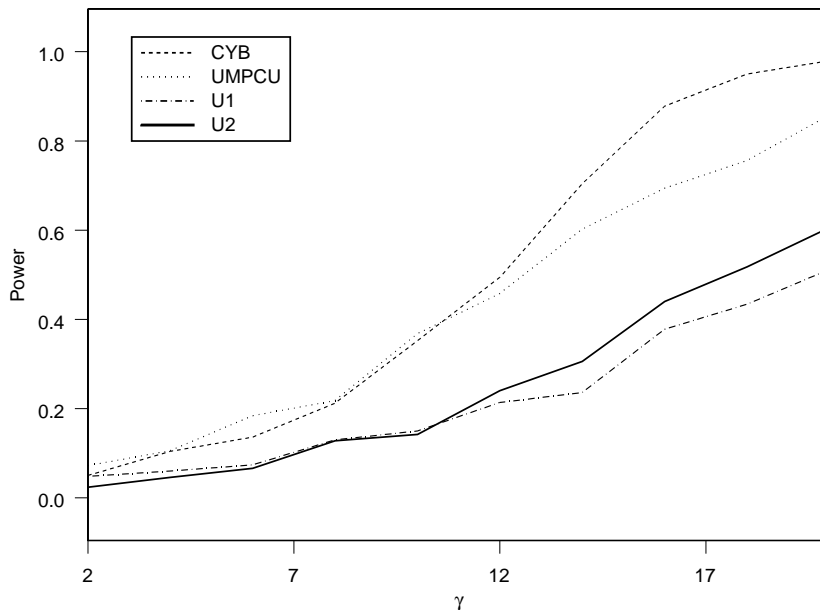
For purposes of comparison with Jansson and Moreira (2006), we set $\beta = \sqrt{1 - \sigma_{ee}^2} \gamma$ and let gamma increase. The power functions for linear alternatives are shown in Figure 1. We notice that the Bonferroni test procedure of Campbell and Yogo (2006) dominates all other tests for linear alternatives. Even though our U statistics are less powerful than the tests based on linear alternatives, they still have power against linear alternatives

For an example of a nonlinear alternative, we let

$$g(x_{t-1}) = \beta x_{t-1}^2,$$

FIGURE 1

Linear Alternatives



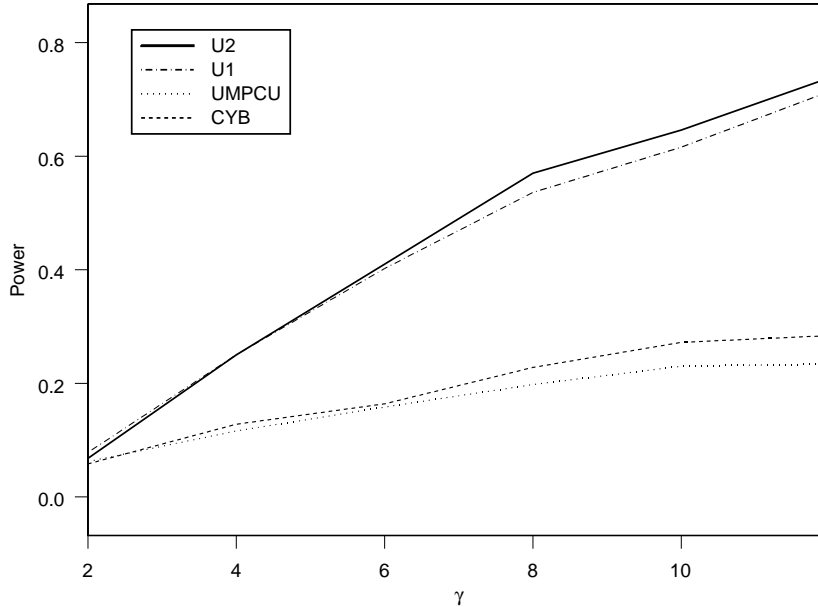
so that the alternative is a quadratic. The power curves are presented in Figure 2. As expected, the U statistics have the highest power. One phenomenon that we found for quadratic alternatives was that the linear based tests (UMPCU and CYB) had power functions that reached a plateau at around 30% and never increased. This power function shows the possible gains from using a U statistic in predictive regression models.

6. EMPIRICAL ILLUSTRATION

The procedures discussed in this paper can be used to test the significance of the regressor x_{t-1} without specifying a functional form. We use a subset of the data discussed in Campbell and Yogo (2006). In particular, we let y_t represent excess monthly returns from the NYSE/AMEX value-weighted index from 1952 to 2002. Moreover, the excess returns are calculated by subtracting the

FIGURE 2

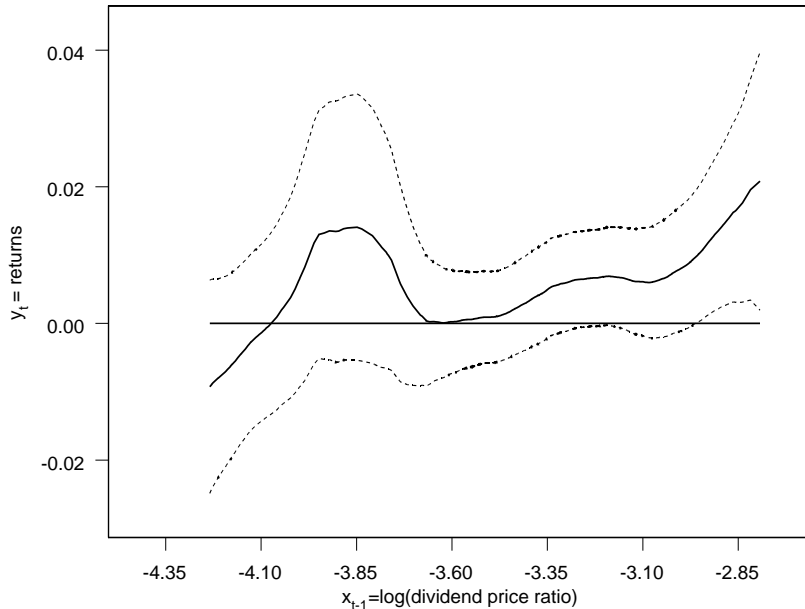
Quadratic Alternatives



one-month T-bill rate for each month. The series x_t represents the dividend-price ratio computed using dividends over the past year divided by price. We have 612 observations. The unit root test of Elliott, Rothenberg, and Stock (1996) is calculated using the modified SIC and Akaike criteria developed in Ng and Perron (2001) to select the number of lags in the test. One lag of x_t is selected by both criteria and the test value is -0.243 which is not significant at the 10% level. Hence, we fail to reject the unit root hypothesis for the log of the dividend-price ratio. This suggests that a standard t-ratio in a regression of y_t on x_{t-1} would not be well approximated by a standard normal distribution.² In light of the persistence in x_{t-1} , we employ the procedure in Campbell and Yogo (2006) using the Q-test. Based on inverting the Q-test, a

²From Table 1, we see that if the null hypothesis of no relationship between y_t and x_{t-1} is true, we would still reject the null hypothesis around 40 % of the time.

FIGURE 3

Estimated $g(x_{t-1})$ 

90% confidence interval for the regression parameter is $(-0.004, 0.010)$ which contains zero. Hence there is no evidence for *linear* predictability of returns using the log of the dividend-price ratio.

To consider the possibility of a nonlinear relationship, we conduct a non-parametric kernel regression of y_t on x_{t-1} . As noted in Bandi (2004), kernel regression can be standardized so that the normal approximation still holds if x_{t-1} has a unit root or near unit root. We tried several bandwidth parameters, and the results were not sensitive within a large range of bandwidths. A representative graph of the estimated function as well as 95 % point-wise confidence bands is shown in Figure 3 using a bandwidth of $h = 2 \times \sigma_x T^{-1/5}$.

There are several features to note about the estimated regression function in Figure 3. Although the estimated function looks highly nonlinear, we notice

that the 95 % point-wise confidence bands contain zero when the log dividend-price ratio is low, so that the null hypothesis of no effect cannot be rejected. However, at higher levels of the log dividend-price ratio, zero is not in the confidence bands, and so one would presume that we can reject the null of no effect of x_{t-1} on y_t for higher levels of x_{t-1} . However, this is not a valid statistical argument. The width of each confidence interval is based on a point-wise argument which means at each point, the confidence interval is valid if we consider the function *at that point in isolation*. However, in this graph, we have estimated the function at 612 points. Therefore, the confidence bands shown in the graph are not confidence bands for the entire function. One adjustment is to use a Bonferroni argument to make the confidence bands wider. Such an adjustment is often extremely conservative, especially when many points of the function are estimated, as in this case. Eubank and Speckman (1993) compare the point-wise bands with the conservative Bonferroni bands to verify the increased width. An alternative approach is to construct asymptotically accurate confidence bands using the approach of Härdle (1989). Such an approach is typically complicated. Moreover, it is unknown if such an approach is valid when the data is persistent to the point of a unit root.

An alternative to correcting the confidence bands for the nonparametric regression function is to simply conduct the U test developed in this paper. We calculate the test and get a value of 0.33. The U test is normally distributed and rejects for large values. Hence, we fail to reject the null hypothesis of no predictability ($g(x_{t-1}) = 0$). In fact, the p-value for a test statistic of 0.33 is 0.38, so that there is very little evidence of predictability. Inference using the U statistic changes the conclusion found using a nonparametric function with point-wise confidence bands.

7. CONCLUSION

In this paper, we have examined the asymptotic distribution theory of non-parametric tests applied to predictive regression models. The most useful feature of the tests is the lack of the so-called “razor’s edge” in asymptotic theory typically associated with unit root models. The limiting distribution for the properly standardized statistic is standard normal if the predictor variable is stationary or if it has a unit root.

A Monte Carlo experiment shows that the test has good size properties and has power against nonlinear alternatives that tests based on linear models tend to miss.

An empirical illustration is used to show the usefulness of the test. In our example, we fail to find evidence for predictability of returns using the log dividend-price ratio as a predictor. The example shows how the test is an important complement to tests based on a linear model and standard non-parametric estimation. In particular, the nonparametric U statistic provides a way to circumvent the currently intractable problem of finding accurate (not conservative) uniform confidence bands for nonparametric regression functions if the regressor has a unit root. In our example, point-wise confidence bands suggest the existence of predictability of returns at very high levels of the dividend-price ratio. However, this is due to the liberal nature of point-wise confidence bands. The U statistic provides a simple way to test the significance of a nonparametric regression function uniformly across all observed data in a predictive regression model.

APPENDIX A

Proof of Theorem 4.1 Use the following definition

$$X_{Tt} = \frac{1}{T^{3/4}\sqrt{h}} \sum_{s=1}^{t-1} K\left(\frac{x_{t-1} - x_s}{h}\right) \epsilon_t \epsilon_s$$

Then X_{Tt} is a martingale and we use the martingale central limit theorem (Theorem 3.2 of Hall and Heyde (1980)) to show our result. There are two conditions that must hold to apply the theorem. Let $\mathcal{F}_t = \sigma(\epsilon_t, e_t, \epsilon_{t-1}, e_{t-1}, \dots)$. To apply the martingale central limit theorem, we must have

$$(A.1) \quad \sum_{t=2}^T E(X_{Tt}^2 1(|X_{Tt}| > \xi_1) | \mathcal{F}_{t-1}) \xrightarrow{p} 0$$

for all $\xi_1 > 0$, and

$$(A.2) \quad \sum_{t=2}^T E(X_{Tt}^2 | \mathcal{F}_{t-1}) \xrightarrow{p} \eta^2$$

where η^2 is an a.s. finite random variable. Note that

$$\begin{aligned} E(X_{Tt}^2 1(|X_{Tt}| > \xi_1) | \mathcal{F}_{t-1}) &= E(X_{Tt}^2 1\left(\frac{|X_{Tt}|}{\xi_1} > 1\right) | \mathcal{F}_{t-1}) \\ &\leq \frac{1}{\xi_1^2} E(X_{Tt}^4 | \mathcal{F}_{t-1}) \end{aligned}$$

Now using Markov's inequality we have

$$P\left(|E(X_{Tt}^4 | \mathcal{F}_{t-1})| \geq \xi_2\right) \leq \frac{1}{\xi_2} E|E(X_{Tt}^4 | \mathcal{F}_{t-1})|$$

Using these inequalities, we have

$$P\left(\left|\sum_{t=2}^T E(X_{Tt}^2 1(|X_{Tt}| > \xi_1) | \mathcal{F}_{t-1})\right| \geq \xi_2\right) \leq \frac{1}{\xi_1^2 \xi_2} \sum_{t=2}^T E(X_{Tt}^4)$$

for all $\xi_1 > 0$, $\xi_2 > 0$. We will show that $\sum_{t=2}^T E(X_{Tt}^4) \rightarrow 0$ which implies that (A.1) holds. Let

$$K\left(\frac{x_{t-1} - x_s}{h}\right) = K_{ts}.$$

We have

$$\begin{aligned}
\sum_{t=2}^T E(X_{Tt}^4) &\sim \frac{1}{T^3 h^2} \sum_{t=2}^T \sum_{s=1}^{t-1} E(K_{ts}^4 \epsilon_t^4 \epsilon_s^4) \\
&+ \frac{1}{T^3 h^2} \sum_{t=3}^T \sum_{s=2}^{t-1} \sum_{r=1}^{s-1} E(K_{ts}^2 K_{tr}^2 \epsilon_t^4 \epsilon_s^2 \epsilon_r^2) \\
&+ \frac{1}{T^3 h^2} \sum_{t=4}^T \sum_{s=3}^{t-1} \sum_{r=2}^{s-1} \sum_{r'=1}^{r-1} E(K_{ts} K_{tr} K_{tr'}^2 \epsilon_t^4 \epsilon_s \epsilon_r \epsilon_{r'}^2) \\
&+ \frac{1}{T^3 h^2} \sum_{t=5}^T \sum_{s=4}^{t-1} \sum_{s'=3}^{s-1} \sum_{r=2}^{s'-1} \sum_{r'=1}^{r-1} E(K_{tr'} K_{tr} K_{ts'} K_{ts} \epsilon_t^4 \epsilon_{r'} \epsilon_r \epsilon_{s'} \epsilon_s) \\
&= B_1 + B_2 + B_3 + B_4
\end{aligned}$$

with $r' < r < s' < s < t - 1$. The dominant term is B_4 . We find the order of B_4 by examining

$$E(K_{tr'} K_{tr} K_{ts'} K_{ts} \epsilon_t^4 \epsilon_{r'} \epsilon_r \epsilon_{s'} \epsilon_s) = \mu_{\epsilon^4} E(K_{tr'} K_{tr} K_{ts'} K_{ts} \epsilon_{r'} \epsilon_r \epsilon_{s'} \epsilon_s)$$

where $\mu_{\epsilon^4} = E(\epsilon^4)$. Since x_{t-1} has a unit root, we have $x_{t-1} - x_s = e_s + e_{s+2} + \dots + e_{t-1}$ and we can write the expectation as

$$\begin{aligned}
(A.3) \quad &E \left[K \left(\frac{e_s + \dots + e_{t-1}}{h} \right) K \left(\frac{e_{r'} + \dots + e_{t-1}}{h} \right) K \left(\frac{e_r + \dots + e_{t-1}}{h} \right) \right. \\
&\quad \left. \times K \left(\frac{e_{s'} + \dots + e_{t-1}}{h} \right) \epsilon_{r'} \epsilon_r \epsilon_{s'} \epsilon_s \right]
\end{aligned}$$

Define the following variables

$$\begin{aligned}
u_1 &= e_{r'} \\
u_2 &= e_{r'+1} + \cdots + e_{r-1} \\
u_3 &= e_r \\
u_4 &= e_r + \cdots + e_{s'-1} \\
u_5 &= e_{s'} \\
u_6 &= e_{s'+1} + \cdots + e_{s-1} \\
u_7 &= e_s \\
u_8 &= e_{s+1} + \cdots + e_{t-1} \\
u_9 &= \epsilon_{r'} \\
u_{10} &= \epsilon_r \\
u_{11} &= \epsilon_{s'} \\
u_{12} &= \epsilon_s
\end{aligned}$$

Let $f_{i,j}$ denote the joint densities of the random variables u_i and u_j and let f_i be the marginal density. Now (A.3) can be written as

$$\begin{aligned}
(A.4) \quad & \int \cdots \int \left[K \left(\frac{u_1 + u_2 + \cdots + u_8}{h} \right) K \left(\frac{u_3 + u_4 + \cdots + u_8}{h} \right) \right. \\
& \times K \left(\frac{u_5 + u_6 + \cdots + u_8}{h} \right) K \left(\frac{u_7 + u_8}{h} \right) u_9 u_{10} u_{11} u_{12} \left. \right] \\
& \times f_{1,9}(u_1, u_9) f_{3,10}(u_3, u_{10}) f_{5,11}(u_5, u_{11}) f_{7,12}(u_7, u_{12}) \\
& \times f_2(u_2) f_4(u_4) f_6(u_6) f_8(u_8) du_1 \dots du_{12}
\end{aligned}$$

We use the change of variables

$$\begin{aligned}
w_1 &= \frac{u_1 + \cdots + u_8}{h} \\
w_2 &= \frac{u_3 + \cdots + u_8}{h} \\
w_3 &= \frac{u_5 + \cdots + u_8}{h} \\
w_4 &= \frac{u_7 + u_8}{h} \\
w_5 &= u_1 \\
w_6 &= u_3 \\
w_7 &= u_5 \\
w_8 &= u_7 \\
w_9 &= u_9 \\
w_{10} &= u_{10} \\
w_{11} &= u_{11} \\
w_{12} &= u_{12}
\end{aligned}$$

which has Jacobian h^4 so that (A.4) can be written as

$$\begin{aligned}
(A.5) \quad & \int \cdots \int \left[K(w_1) K(w_2) K(w_3) K(w_4) w_9 w_{10} w_{11} w_{12} \right] \\
& \times f_{1,9}(w_5, w_9) f_{3,10}(w_6, w_{10}) f_{5,11}(w_7, w_{11}) f_{7,12}(w_8, w_{12}) \\
& \times f_2(hw_1 - hw_2 - w_5) f_4(hw_2 - hw_3 - w_6) \\
& \times f_6(hw_3 - hw_4 - w_7) f_8(hw_4 - w_8) h^4 dw_1 \dots dw_{12}
\end{aligned}$$

From Assumption 4, the density functions f_2, f_4, f_6 , and f_8 are bounded by constants times $(\sqrt{r - r'})^{-1}, (\sqrt{s' - r})^{-1}, (\sqrt{s - s'})^{-1}$, and $(\sqrt{t - s})^{-1}$ respectively. Hence, the term B_4 is of order

$$\frac{1}{T^3 h^2} \sum_{t=5}^T \sum_{s=4}^{t-1} \sum_{s'=3}^{s-1} \sum_{r=2}^{s'-1} \sum_{r'=1}^{r-1} (\sqrt{r - r'} \sqrt{s' - r} \sqrt{s - s'} \sqrt{t - s})^{-1} h^4 \sim \frac{1}{T^3 h^2} T (\sqrt{T})^4 h^4$$

Hence, $B_4 \rightarrow 0$ since $h \rightarrow 0$. Similarly, B_1, B_2 , and B_3 can be shown to converge to zero, proving that (A.1) holds.

Now, we prove that (A.2) holds.

$$\begin{aligned} \sum_{t=2}^T E(X_{Tt}^2 | \mathcal{F}_{t-1}) &= \frac{1}{T^{3/2}h} \sum_{t=2}^T \sum_{s=1}^{t-1} K_{ts}^2 \epsilon_s^2 \sigma_\epsilon^2 + \frac{1}{T^{3/2}h} \sum_{t=2}^T \sum_{s \neq s'}^{t-1} K_{ts} K_{ts'} \epsilon_s \epsilon_{s'} \sigma_\epsilon^2 \\ &= C_1 + C_2 \end{aligned}$$

Consider the term C_1 . Define $M(a, b)$ as $B(b) - B(a)$ where $B(\cdot)$ is a Brownian motion with variance σ_ϵ^2 . Following Phillips and Park (1998), we can write

$$\begin{aligned} \frac{1}{T^{3/2}h} \sum_{t=2}^T \sum_{s=1}^{t-1} K_{ts}^2 &= \frac{1}{T^{3/2}h} \sum_{t=2}^T \sum_{s=1}^{t-1} K^2 \left(\frac{e_s + e_{s+1} + \dots + e_{t-1}}{h} \right) \\ &= \frac{\sqrt{T}}{h} \int_0^1 \int_0^b K^2 \left(\frac{\sqrt{T}}{h} M(a, b) \right) da db + o_{a.s.}(1) \\ &= \int_0^1 \left(\int_{-\infty}^{\infty} K^2(s) ds \right) L_M(b, 0) db + o_{a.s.}(1) \end{aligned}$$

where $L_M(b, 0)$ is the chronological local time of the process M at the origin.

We can then apply Theorem 3.3 of Hansen (1992) to get

$$\frac{1}{T^{3/2}h} \sum_{t=2}^T \sum_{s=1}^{t-1} K_{ts}^2 (\epsilon_s^2 \sigma_\epsilon^2 - \sigma_\epsilon^4) \xrightarrow{p} 0$$

Now consider C_2 . We have

$$E(C_2^2) = \frac{1}{T^3 h^2} \sum_{t=2}^T \sum_{t'=2}^T \sum_{s \neq s'}^{t-1} \sum_{r \neq r'}^{t'-1} \sum_{t''=1}^{t-1} \sum_{t'''=1}^{t'-1} E(K_{ts} K_{ts'} K_{t'r} K_{t'r'} \epsilon_s \epsilon_{s'} \epsilon_r \epsilon_{r'} \sigma_\epsilon^4)$$

There are several cases to examine for the indices. Consider the case were $r \neq r' \neq s \neq s'$. Since the variables are jointly elliptically distributed, $E(\epsilon_s | e_s)$ is a linear function of e_s . Therefore,

$$E(K_{ts} K_{ts'} K_{t'r} K_{t'r'} \epsilon_s \epsilon_{s'} \epsilon_r \epsilon_{r'}) \sim E(K_{ts} K_{ts'} K_{t'r} K_{t'r'} e_s e_{s'} e_r e_{r'})$$

Using a similar change of variables as before, we have

$$\begin{aligned} (A.6) \quad & \int \dots \int \left[K(w_1) K(w_2) K(w_3) K(w_4) w_5 w_6 w_7 w_8 \right] \\ & \times f_1(w_5) f_3(w_6) f_5(w_7) f_7(w_8) \\ & \times f_2(hw_1 - hw_2 - w_5) f_4(hw_2 - hw_3 - w_6) \\ & \times f_6(hw_3 - hw_4 - w_7) f_8(hw_4 - w_8) h^4 dw_1 \dots dw_8 \end{aligned}$$

We can take a mean value expansion of f_2 around $-w_5$ to get

$$f_2(hw_1 - hw_2 - w_5) = f_2(-w_5) + f_2'(c_1)(hw_1 - hw_2)$$

where c_1 is a linear combination of $hw_1 - hw_2$ and w_5 . However, by Assumption 4, $f_2'(c_1)$ is bounded by a c_1 times $(r - r')^{-3/2}$ since the variance of u_2 is of order $(r - r')$. Then (A.6) becomes

$$\begin{aligned} & \int \cdots \int \left[K(w_1) K(w_2) K(w_3) K(w_4) w_5 w_6 w_7 w_8 \right] \\ & \quad \times f_1(w_5) f_3(w_6) f_5(w_7) f_7(w_8) \\ & \quad \times f_2(-w_5) f_4(hw_2 - hw_3 - w_6) \\ & \quad \times f_6(hw_3 - hw_4 - w_7) f_8(hw_4 - w_8) h^4 dw_1 \dots dw_8 \\ + & \int \cdots \int \left[K(w_1) K(w_2) K(w_3) K(w_4) w_5 w_6 w_7 w_8 \right] \\ & \quad \times f_1(w_5) f_3(w_6) f_5(w_7) f_7(w_8) \\ & \quad \times f_2'(c_1)(hw_1 - hw_2) f_4(hw_2 - hw_3 - w_6) \\ & \quad \times f_6(hw_3 - hw_4 - w_7) f_8(hw_4 - w_8) h^4 dw_1 \dots dw_8 \\ & = D_1 + D_2. \end{aligned}$$

Since $f_2(-w_5)w_5$ is symmetric about the origin, $\int f_2(-w_5)w_5 f_1(w_5)dw_5 = 0$, so that $D_1 = 0$. The order of D_2 is given as $(r - r')^{-3/2}(s' - r)^{-1/2}(s - s')^{-1/2}(t - s)^{-1/2}h^4$. Hence, for the case when $r \neq r' \neq s \neq s'$, $E(C_2^2)$ is of order

$$\frac{1}{T^3 h^2} \sum_{t=2}^T \sum_{t'=2}^T \sum_{s \neq s'}^{t-1} \sum_{r \neq r'}^{t-1} \sum_{r' \neq r'}^{t'-1} (r - r')^{-3/2} (s' - r)^{-1/2} (s - s')^{-1/2} (t - s)^{-1/2} h^4 = h^2$$

The other cases are similar so that the order of $E(C_2^2) = h^2$ and hence C_2 converges to zero.

REFERENCES

- BANDI, F. M. (2004): "On the Persistence and Nonparametric Estimation with an Application to Stock Return Predictability," Graduate School of Business, University of Chicago.
- BERK, J. (1997): "Necessary Conditions for the CAPM," *Journal of Economic Theory*, 73, 245–257.

- CAMPBELL, J. Y., AND M. YOGO (2006): “Efficient Tests of Stock Return Predictability,” *Journal of Financial Economics*, 81, 27–60.
- CAVENAGH, C., G. ELLIOTT, AND J. STOCK (1995): “Inference in Models with Nearly Integrated Regressors,” *Econometric Theory*, 11, 1131–1147.
- DICKEY, D. A., AND W. A. FULLER (1979): “Distribution of the Estimators for Autoregressive Time Series with a Unit Root,” *Journal of the American Statistical Society*, 74, 427–431.
- EUBANK, R., AND P. SPECKMAN (1993): “Confidence Bands in Nonparametric Regression,” *Journal of the American Statistical Association*, 88, 1287–1300.
- FAN, Y., AND Q. LI (1996): “Consistent Model Specification Tests: Omitted Variables and Semi-parametric Functional Forms,” *Econometrica*, 64, 865–890.
- (1999): “Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing,” *Journal of Nonparametric Statistics*, 10, 245–271.
- HALL, P., AND C. HEYDE (1980): *Martingale Limit Theory and its Application*. New York: Academic Press.
- HANSEN, B. E. (1992): “Convergence to Stochastic Integrals for Dependent Heterogeneous Processes,” *Econometric Theory*, 8, 489–500.
- HÄRDLE, W. (1989a): *Applied Nonparametric Regression*. Cambridge: Cambridge University Press.
- (1989b): “Asymptotic Maximal Deviation of M-Smothers,” *Journal of Multivariate Analysis*, 29, 163–179.
- HODGSON, D. J., O. LINTON, AND K. VORKINK (2002): “Testing the Capital Asset Pricing Model Efficiently Under Elliptical Symmetry: A Semiparametric Approach,” *Journal of Applied Econometrics*, 17, 617–639.
- JANSSON, M. (2006): “Semiparametric Power Envelopes for Tests of the Unit Root Hypothesis,” Department of Economics, University of California, Berkeley.
- JANSSON, M., AND M. J. MOREIRA (2006): “Optimal Inference in Regression Models with Nearly Integrated Regressors,” *Econometrica*, 74, 681–714.
- LEWELLEN, J. (2004): “Predicting Returns with Financial Ratios,” *Journal of Financial Economics*, 74, 209–235.

- LI, Q. (1999): “Consistent Model Specification Tests for Time Series Econometric Models,” *Journal of Econometrics*, 92, 101–148.
- NG, S., AND P. PERRON (2001): “Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power,” *Econometrica*, 69, 1519–1554.
- OWEN, J., AND R. RABINOVITCH (1983): “On the Class of Elliptical Distributions and Their Applications to the Theory of Portfolio Choice,” *Journal of Finance*, 38, 745–752.
- PHILLIPS, P. C. B., AND J. Y. PARK (1998): “Nonstationary Density Estimation and Kernel Autoregression,” Cowles Foundation, Yale University.
- SHIN, D. W., AND B. S. SO (1999): “Unit Root Tests Based on Adaptive Maximum Likelihood Estimation,” *Econometric Theory*, 15, 1–23.
- STAMBAUGH, R. F. (1999): “Predictive Regressions,” *Journal of Financial Economics*, 54, 375–421.
- ZHENG, J. X. (1996): “A Consistent Test of Functional Form via Nonparametric Estimation Techniques,” *Journal of Econometrics*, 75, 263–290.