Long-Horizon Regressions when the Predictor is Slowly Varying[1]

Roger Moon
USC[2]

Antonio Rubia
University of Alicante and UCSD[3]

Rossen Valkanov
UCSD[4]

November 23, 2005

[1]We thank Alberto Plazzi, Pedro Santa-Clara, Walter Torous, Michael Wolf and seminar participants at University of Navarra and CORE for useful comments and suggestions. All remaining errors are our own.

[2]Los Angeles, CA 90089, phone: (213) 740-2108, e-mail: moonr@usc.edu.

[3]Alicante, CP 03080. Phone (34) 965 90 36 21. e-mail: antonio.rubia@ua.es; and La Jolla, CA 92093-0093. Phone: (858) 534-0895.

[4]Corresponding author. La Jolla, CA 92093-0093. Phone: (858) 534-0898. Email: rvalkanov@ucsd.edu

Long-Horizon Regressions when the Predictor is Slowly Varying

Abstract: Predictive stock return regressions have two distinctive characteristics: (i) the predictor on the right-hand side is persistent and its variance is orders of magnitude smaller than the variance of returns; (ii) the left-hand side variable is a long-horizon return constructed from overlapping sums of short-horizon returns. We offer a new model for the predictor that parsimoniously captures and links its persistence and small variance. We then use two asymptotic approaches to analyze the properties of long-horizon regressions. The approaches differ in their treatment of the overlap. One of the asymptotics has previously been analyzed with other data generating processes, while the second one is novel. We find that under both asymptotics, least-squares estimators may not be consistent, their t-statistics diverge, and the R^2 is not an adequate goodness-of-fit measure. Interestingly, a re-scaled version of the t-statistic is consistent under both long-horizon approximations and is suitable for testing predictability in long-horizon regressions. A Monte Carlo analysis of the finite-sample properties of the re-scaled t-statistic reveals that both approximations are accurate even for small sample sizes. We apply these results to test for predictability in returns of real estate investment trusts (REITs) which have come into existence only since the early 1970s and for which a reliable predictability test is crucial given the small dataset.
JEL Classification: C22, G14, G12.
Keywords: return predictability, long-horizon regressions, near unit root, local-to-zero variance.

1 Introduction

Return predictability has a central place in empirical finance due to its strong implications for asset pricing theory and practical portfolio management. Despite the considerable body of literature on this issue, predictability remains as an open question and is at the forefront of empirical asset pricing. Campbell (1987), Campbell and Shiller (1988, 1989), Fama and French (1988), Fama and Schwert (1977), Hodrick (1992) and several others find evidence of predictability, while Ang and Bekaert (2003), Bossaerts and Hillion (1999), Ferson, Sarkissian and Simin (2003), Goetzmann and Jorion (1993), and Torous, Valkanov and Yan (2004) do not. Baker and Wurgler (2000), Campbell and Yogo (2003), Lanne (2002), Lettau and Ludvigson (2001), Lewellen (2004), and Valkanov (2003) offer new contributions on this topic.[1] The overwhelming majority of these studies are conducted using time-series regressions that have two distinctive characteristics. First, the forecasting variable is a highly-persistent process with a variance orders of magnitude smaller than the variance of returns. Second, the dependent variable is a long-horizon return constructed from overlapping sums of short-horizon returns. The conflicting findings in this literature are mostly due to differences in the approach chosen to deal with these statistical issues. The divergence of results also suggests that the properties of long-horizon regressions on persistent, slowly varying predictive variables need further studying.

In this paper, we offer new results on the properties of predictive regressions by accounting for these key statistical features, and contribute to the return predictability literature in several ways. First, we introduce a new process that explicitly links the persistence and the small variance of the predictor. This process allows us to capture more accurately the statistical properties of long-horizon regressions. In particular, while several recent papers have demonstrated the importance of persistent forecastors in long-horizon regressions (see, among others, Campbell and Yogo 2003; Lewellen 2004; Stambaugh 1999; and Torous *et al.* 2004), the related studies do not incorporate the observations in Fama and French (1988), Shiller (1981) and Summers (1986), who point out that predictive variables are systematically less volatile than

[1] The literature on return predictability is truly voluminous. Our apologies to everyone whose paper we omit to cite here or below.

returns.[2] Furthermore, it is well-known that a link ought to exist between the small variability in the conditioning variable and its persistence. For instance, Campbell, Lo, and MacKinlay (1997) argue that the more persistent is the forecasting variable, the smaller its variance must be (and vice versa). Otherwise, a persistent predictor exhibiting a large variance would imply large return predictability even at short-horizons, which is counter-factual and not economically appealing. The explicit modelling of the small variability and its link to the persistence of the predictor has been neglected in the empirical finance and econometrics literature, because it is often argued that a small-variance does not affect the asymptotic properties of statistics in the regressions. We show, however, that the small-sample properties of long-horizon predictive regressions are better approximated when this feature of the data is captured explicitly and, more importantly, that it accounts for several empirical findings in the predictability literature.

Second, we use two alternative approaches to model the degree of overlap in long-horizon regressions and thus provide an exhaustive analysis of their statistical properties. Lanne (2002), Richardson and Stock (1989), and Valkanov (2003) show that if the return horizon, denoted by K, is a non-trivial fraction of the sample size T, then the small-sample properties of the test statistics of interest can accurately be approximated using asymptotic results that model the ratio K/T as converging to a fixed constant κ $(0 < \kappa < 1)$. We apply this asymptotic approach, which allows the overlap to increase at the same rate as the sample size, with the new data generating process to derive the properties of long-horizon forecasting regressions. In addition, we consider a new asymptotic approximation, in which the ratio K/T converges to zero at various rates. Under this approach, the degree of overlap is allowed to increase at a rate slower than the sample size. Since $0 < K < T$, the analysis under both approximations captures all the relevant possibilities for empirical purposes with the aim of addressing the properties of long-horizon regressions as a function of the degree of overlap. Furthermore, since both asymptotic approaches are only approximations designed to understand the small-sample properties of

[2] Ferson *et al.* (2003), for instance, provide summary statistics of the most widely used predictors of returns. Table 1 in their paper (pp. 1396) illustrates the persistence and small variability of these predictors. Their first-order autocorrelation parameter is close to one and their variance is 50 to 100 times smaller than the variance of returns.

long-horizon predictive regressions, there are no a priori reasons to believe that one is better than the other. In considering both approaches, we not only provide an exhaustive analysis on modeling long-horizons regressions, but we also compare how these two approximations differ and if there are obvious advantages of using one rather than the other.

We find that both asymptotic methods yield remarkably similar results regarding the properties of long-horizon predictive regressions. More specifically, under the general setting of both approaches the least-squares estimator is not consistent. Moreover, the OLS t-statistic tends to increase with the overlap and cannot provide a consistent testing procedure under the null of no predictability. We show that it converges to a well-defined distribution only after it is scaled by $\sqrt{K}$. Finally, the regression $\mathcal{R}_K^2$ is not an adequate goodness-of-fit measure for small samples under either approximation. For certain values of the nuisance parameters, the asymptotic distributions are virtually indistinguishable in small samples. Moreover, we use Monte Carlo simulations to show that the asymptotic approximations are similar to each other for various values of the nuisance parameters. We also find that both asymptotic approximations are accurate in capturing the small sample properties of long-horizon regressions with persistent and slowly varying predictors. Even in small samples with significant overlap, we obtain good results.

Our asymptotic findings provide a remarkably clear and unified guidance on how to conduct inference in long-horizon regression. Firstly, no matter which theoretic approach we use, the natural test for long-horizon predictability is a rescaled OLS t-statistic, namely $t/\sqrt{K}$. It converges to a well-defined distribution under both asymptotic approaches. Secondly, the distribution of the tests will depend on nuisance parameters (including, at least, the correlation between contemporaneous shocks in returns and predictors) which, in real applications, will be estimated consistently from the data. Ultimately, these distributions must be simulated in order to conduct inference. Thirdly, our results imply that long-horizon regressions do not generally increase the power of predictability tests. All three of these implications are verified with Monte Carlo simulations.

Using our new results, we investigate whether the returns of real estate investment trusts (REITs) are predicted by their dividend yields. We focus on REIT companies because there are reasons to believe that predictability might be more evident for these assets. Indeed, REITs must by law pay out a large fraction (at least 90%) of their taxable income to shareholders in the form of dividends. In contrast, common stock dividends are paid at the discretion of the firm's management and there is ample evidence that they are actively smoothed, the product of managers catering to investors' demand for dividends, or the result of managers' reaction to perceived mispricings (Shefrin and Statman 1984; Stein 1996; Baker and Wurgler 2004). Thus, it has been argued that common stock dividends may not accurately reflect changing investment opportunities surrounding a firm. Moreover, the dividends of non-REIT companies are also subject to long-term trends such as the recent low propensity of firms to pay dividends (Fama and French 2001) which further results in them being even more persistent and even less variable. Therefore, there are natural advantages in addressing predictability on REIT dividends. A drawback of using REIT to test for predictability is that they are relatively new assets. We only have data since the early 1970s and the sector has matured and changed dramatically over the last fifteen years. Hence, an analysis on the basis of a theory that provides a robust approach for small samples is especially important. We find that while the dividend yield has some predictive power for the REIT stock returns, the evidence is not as overwhelming as standard inference suggests. These results confirm the findings in earlier studies and underline the importance of using accurate inference methods.

Torous and Valkanov (2003) offer an alternative model of persistent predictors with small variance. However, our analysis differs from their in several respects. First, they do not provide a link between the persistence of the predictor and its variability. Instead, they introduce two parameters that need to be calibrated. In our model, the persistence and the small variance are connected and captured by a single parameter. This is appealing from an economic perspective, as discussed above, and also on the grounds of econometric parsimony. Moreover, Torous and Valkanov (2003) do not analyze the importance of the small variability assumptions for long horizon regressions, where predictability is most often tested. They focus exclusively on short-horizon, one-period regressions and on their in-sample and out-of-sample properties.

The paper is organized as follows. Section two states the data generating process for returns in our analysis and discusses its implications. In Section three, we derive the main results for long-horizon regressions. Section four investigates the small-sample properties of the statistics through Monte Carlo experimentation. Section five discusses the predictability of REIT stocks. Finally, Section six summarizes and concludes.

2 A Model for Returns and Dividend Yields

The focus of this paper is on the statistical properties of long-horizon predictive regressors. We start the analysis by specifying the stochastic behavior of the one-period log returns time-series, denoted by $\{r_t, t \geq 1\}$, and of the forecasting variable, $\{x_t, t \geq 1\}$, which we take to be the log of the dividend yield. This variable has been the most common predictor employed in empirical tests. The relation between the one period returns and the predictor is specified as

$$r_{t+1} = \mu_r + \beta x_t + u_{t+1} \tag{1}$$

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

Campbell (2001), Campbell and Shiller (1988), Cochrane (2001), Mankiw and Shapiro (1986), Nelson and Kim (1993), Stambaugh (1986, 1999) and many other use this system as a starting point in their analysis. The null hypothesis of interest is the no-forecastability of returns, or that the one-period-ahead returns r_{t+1} cannot be predicted by a de-meaned forecaster x_t. In a more general setting, x_t is given by the set of historical information up to time t, say $\mathcal{H}_t$. This set can include lagged returns and any other (predetermined) variable, such as the term spread, the default spread, various interest rates, inflation, and consumption-wealth ratios (Campbell 1987; Fama and French 1988; Lettau and Ludvigson 2001). With the exception of lagged returns, all these predictors are persistent and have small variance in comparison with returns. Hence, our treatment generalizes to these variables. It is argued that under the alternative of predictability, or $E\left(r_{t+1} | \mathcal{H}_t\right) = \mu_r + \beta x_t$ for $\beta > 0$, the most likely source of variability originates from changes in investors' perception of risk, which are eventually reflected in a time-varying risk premium. Thus, the dividend yield is widely considered to be a noisy proxy for time-varying expected returns (Campbell and Shiller 1988; Fama and French 1988).

Equation (1) is the predictive regression, while equation (2) parsimoniously models the dynamics of the dividend yield as an AR(1) process. The persistence of the dividend yield is captured by ρ, where

$$\rho = 1 - c/T \tag{3}$$

and T is the sample size. In this specification, ρ is in a $1/T$ neighborhood to unity and c is a non-centrality parameter. This local-to-unity (or near-integrated) specification has been used extensively to model persistent but stationary processes, and it is well-known in the non-stationary literature (Cavanagh, Elliott and Stock 1998; Elliott 1998; Phillips 1987; Stock 1991). We consider c in the region $c > 0$ in order to rule out integrated or explosive processes. This restriction is necessary, because non-stationary predictors are difficult to justify on economic grounds[3]. The vector of error terms $(u_t, \varepsilon_t)'$ is generally assumed to be independent across t but correlated contemporaneously (Campbell 2001; Stambaugh 1999). We shall denote $E\left(u_t^2\right) = \sigma_u^2$, $E\left(\varepsilon_t^2\right) = \sigma_\varepsilon^2$, and $E\left(u_t \varepsilon_t\right) = \gamma \sigma_u \sigma_\varepsilon$, where $\gamma \in (-1, 1)$.

The predictive variable x_t is not only characteristically persistent, but its unconditional variance is small in relation to the variance of unexpected returns, σ_u^2. In other words, while expected returns might be time-varying, or $E\left(r_{t+1} | \mathcal{H}_t\right) = \mu_r + \beta x_t$ for $\beta > 0$, their variability must be a small fraction of the total variance of r_{t+1}. Otherwise, short-horizon returns would easily be forecastable. To model the small variability of the predictor, we assume that

$$\varepsilon_t = \sigma_u \sqrt{1 - \rho} e_t \tag{4}$$

where e_t is a stationary error process with zero mean and variance σ_e^2. Without loss of generality, we set $\sigma_e = 1$, because this scale parameter does not play a major role in the main conclusions of our analysis. From the restriction (4) it follows that $E\left(\varepsilon_t^2\right) \equiv \sigma_u^2 c/T$ and, hence, the unconditional variance of each shock to x_t is local-to-zero.[4]

[3] Persistence in the dividend yield has received a considerable attention in the recent forecastability literature; see, among others, Campbell *et al.* (1997), Stambaugh (1999), and Ferson *et al.* (2003). Although dividend yields are highly persistent, the presence of an exact unit root implies strong statistical features that are hard to justify in practice (see a discussion in Lewellen, 2004).

[4] Since the variance σ_ε^2 depends on T, the system generated from (1) and (2) with the restriction (4) formally constitutes a triangular array $\left\{r_{T,t}, x_{T,t}, \sigma_{\varepsilon T}^2\right\}$. Since the triangular array notation is not essential, for simplicity the subscript T is suppressed from now onward.

Specification (4) is intended to capture the fact that in finite samples the volatility of the predictor is orders of magnitude smaller than the variability of the returns. The parameterization has several appealing features. First, it links the persistence and the small-variability of x_t. The more persistent is the conditioning variable, the smaller should be the variance of ε_t. Otherwise, fluctuations in x_t will have large effects on returns.[5] Second, for the large-sample analysis that will follow, the small variance of ε_t makes x_t to be (asymptotically) of the same stochastic order as u_t. Third, the small variability of ε_t is modelled without additional nuisance parameters. Torous and Valkanov (2003) propose an alternative way of modeling the small variance of ε_t based on an additional parameter that controls the rate at which the variance of ε_t converges to zero. However, that approach does not link the persistence of the predictor to its small variability. Moreover, the practical implementation of their model requires the calibration of an additional parameter. Finally, the different rates of convergence add another level of complexity to the asymptotic results.

We summarize the properties of the predictor in the following two assumptions.

Assumption 1 *The dynamics of the predicting variable follow a near-integrated process given by* (2) *and* (3) *for some fixed initial condition* $x_0 = 0$.

Assumption 2 *The error term* $\eta_t = (u_t, \varepsilon_t)'$ *is independent and identically distributed over time with covariance matrix*

$$\Sigma_\eta = \begin{pmatrix} \sigma_u^2 & \gamma\sigma_\varepsilon\sigma_u \\ \gamma\sigma_\varepsilon\sigma_u & \sigma_\varepsilon^2 \end{pmatrix} \tag{5}$$

for a fixed correlation parameter $-1 < \gamma < 1$ *and* σ_ε^2 *follows* (4).

To understand the importance of the local-to-zero variance specification, consider the local-to-unity predictor (2)-(3) *without* assumption (4) and denote $\lambda_r = \left(e^{2rc} - 1\right)$. Then, for any $t = [Tr]$, with $r \in (0, 1]$, it can be shown that $Var\left(x_{[Tr]}\right) = \sigma_\varepsilon^2 \lambda_r / 2c$. As the process becomes more persistent we see that $\lim_{c \to 0} Var\left(x_{[Tr]}\right) = \sigma_\varepsilon^2 r$. In contrast, under specification (4),

[5] An intuitive way of viewing this restriction is to consider that the variance of the process is some small fraction of the return variability. In our approach, this is achieved by scaling the variance with a factor that depends on the sample size.

$\lim_{c\to 0} Var\left(x_{[Tr]}\right) = \lim_{c\to 0}\sigma_\varepsilon^2\lambda_r/2 = 0$, where, strictly speaking, σ_ε^2 depends on T. We omit this additional subscript for the sake of simplicity. Hence, even for highly-persistent processes, the variance of the predictor will be small, which is the empirical feature largely observed in practice. Moreover, it is worth noting that this property has further implications on the predictive regression tests. Under specification (4), the signal-to-noise ratio in the one-period predictive regression is given by,

$$\left[\frac{\beta^2 Var\left(x_{[Tr]}\right)}{Var(u_{t+1})}\right] \sim \left(e^{2rc}-1\right)\beta^2 c/2 \tag{6}$$

which, clearly, vanishes as $c \to 0$ even if $\beta \neq 0$. In other words, it is difficult to find a predictive relationship through the regression analysis, because the useful information conveyed by the signal observable in the predictor is masked by the greater volatility of the unexpected return. This is largely consistent with the arguments in the predictive literature. In the next section, we use the predictive process defined in Assumptions (1)-(2) to derive the properties of long-horizon regressions.

3 Inference in Long-Horizon Regressions

An implication of Campbell and Shiller's (1988) dynamic Gordon growth model is that the dividend price ratio may forecast long-horizon returns. More precisely, it is argued that long-horizon returns are better at proxying for time-varying expected returns than short-horizon returns, and that long-horizon tests have better power at detecting predictable fluctuations in expected returns. The log-linearized Campbell and Shiller (1988) framework offers a very convenient linear relation between a state variable x_t and future returns. Hence, most of the research on predicting long-horizon returns has been conducted with linear long-horizon regressions.

The long-horizon regression for the K-horizon return, $K > 1$, is simply given by the linear regression model,

$$R_{K,t+1} = \alpha_K + \beta_K x_t + U_{K,t+1} \tag{7}$$

where the dependent variable $R_{K,t+1} = \sum_{j=0}^{K-1} r_{t+1+j}$ is the long-horizon continuously compounded return. The main hypothesis of interest is that of not predictive relation between $R_{K,t+1}$ and x_t, that is, $\mathrm{H}_0 : \beta_K = 0$. We shall focus carefully on the properties of the relevant statistics from (7) under this hypothesis. Consequently, our interest is to understand the properties of the estimated OLS slope coefficient $\hat{\beta}_K$ from (7),

$$\hat{\beta}_K = \frac{\sum_{t=1}^{T-K} R_{K,t+1} (x_t - \bar{x})}{\sum_{t=1}^{T-K} (x_t - \bar{x})^2} \tag{8}$$

and its OLS t-statistic,

$$t = \frac{\sum_{t=1}^{T-K} R_{K,t+1} (x_t - \bar{x})}{\hat{\sigma}_K \left(\sum_{t=1}^{T-K} (x_t - \bar{x})^2 \right)^{1/2}}. \tag{9}$$

We shall also consider the $\mathcal{R}_K^2$ from the regression,

$$\mathcal{R}_K^2 = 1 - \frac{\sum_{t=1}^{T-K} \left(U_{K,t+1}^2 - \bar{U}_{K,t+1}^2 \right)}{\sum_{t=1}^{T-K} \left(R_{K,t+1} - \bar{R}_{K,t+1} \right)^2} \tag{10}$$

under the hypothesis of no-predictability, where $\hat{\sigma}_K^2 = \frac{1}{T-K} \sum_{t=1}^{T-K} \hat{U}_{K,t+1}^2$, $\hat{U}_{K,t+1} = R_{K,t+1} - \hat{\alpha}_K - \hat{\beta}_K x_t$ with $\hat{\alpha}_K$ and $\hat{\beta}_K$ being the OLS estimators of α_K and β_K, and $\bar{U}_{K,t+1}^2$ and $\bar{R}_{K,t+1}$ denoting the average values of $U_{K,t+1}^2$ and $R_{K,t+1}$, respectively.

Assessing predictability in multiperiod regressions implies invariably overlapping returns, owing to the lack of enough independent observations. Since the sample size is fixed, the larger is the return horizon K, the larger is the relative degree of overlap. This property makes the return horizon K a key variable when conducting inference in long-horizon regressions, as the degree of overlap can heavily influence the properties and the limiting distributions of the inference statistics. The analysis of overlapping long-horizon regressions, be in asymptotic or finite sample, must invariably consider the overlap K in relation to the sample size T.

We analyze two different asymptotic approaches to model the properties of long-horizon regressions with overlapping observations, which differ only in their treatment of the overlap K in relation to the sample size. First, we consider an approximation in which K is allowed to diverge with the sample size, but at a slower rate. This includes the trivial case in which K is a fixed value, and the case in which K is bounded by T^{α}, $\alpha \in (0,1)$. Alternatively, we consider an approximation in which $K \sim \kappa T$ for $\kappa \in (0,1)$, so that the overlap grows linearly with the sample size. In both theories the overlap is an increasing function of the sample size, which is a simple and convenient modelling device. The main aim is to prevent the overlap from decreasing in importance as $T \to \infty$, since, otherwise, there would be no difference between a long- and a short-horizon regression.

Ultimately, asymptotic analyses are only useful for practical purposes if they help us understand the properties of $\hat{\beta}_K, t$, and $\mathcal{R}_K^2$ and if they successfully approximate the small-sample distributions of these statistics. Empirical tests of long-horizon predictability with overlapping observations usually exhibit positive values for $\hat{\beta}_K$ coupled with large t-statistics and impressive $\mathcal{R}_K^2$. Moreover, all these statistics grow with the return horizon K. If taken at face value, these results suggest strong predictive ability of the dividend yield at long-horizons. Yet, as discussed previously, long-horizon regressions imply a collection of relevant statistical issues that cannot be ignored. We use our asymptotic approximations to understand the properties of these statistics.

The main tool in our analytic work is the Functional Central Limit Theorem (FCLT), which we will use to analyze the asymptotic behavior of these statistics, their statistical properties as a function of the overlap, and the effects of nuisance parameters. Note that small-sample results typically depend on strong (likely unrealistic) assumptions about the distribution of the data and, furthermore, cannot be universally obtained for complex settings, as the one addressed here.[6] Since weakly convergence holds very fast under some sample characteristics, the limit distributions could provide good approximations even for small sample sizes.

In what follows, $\Rightarrow$ and $\to_p$ denotes weak convergence in distribution and convergence in probability, respectively. We use '$\overset{d}{=}$' to denote equality in distribution. The conventional notation $o(1)$ $(o_p(1))$ is used to represent a series of numbers (random numbers) converging to zero (in probability). $O(1)$, $(O_p(1))$ denotes a series of numbers (random numbers) that are bounded (in probability). The limiting forms of the test statistics are expressed as functionals of Brownian motion processes. The following Lemma follows from Assumptions (1)-(2) and Herrndorf (1984) and Phillips (1987).

[6] Since the predictor is not strictly exogenous –only predetermined– the properties of the statistics can only be formally addressed under the asymptotic theory.

Lemma 3.1 *Denote the demeaned process* $\tilde{x}_t = x_t - T^{-1}\sum_{t=1}^{T} x_t$. *Under Assumptions 1-2, the following result holds as* $T \to \infty$,

$$\left(\frac{1}{\sigma_u}\sum_{t=1}^{[Tr]}\varepsilon_t, \frac{1}{\sigma_u\sqrt{T}}\sum_{t=1}^{[Tr]}u_t, \frac{1}{\sigma_u}\tilde{x}_{[Tr]}\right)' \Rightarrow \left(\sqrt{c}W_\varepsilon(r), W_u(r), \sqrt{c}\tilde{J}_c(r)\right)'; \quad r \in [0,1]$$

in $D[0,1] \times D[0,1] \times D[0,1]$, *where* $[W_\varepsilon(r), W_u(r)]'$ *is a vector of standard Wiener processes with correlation* γ, *and* $\tilde{J}_c(r)$ *is a demeaned Ornstein-Uhlenbeck diffusion process on* $W_\varepsilon(r)$.

3.1 Asymptotic Properties: K/T$\to 0$

In this asymptotic framework, the overlap K in predictive regressions grows at a rate slower than the sample size. This assumption allows us to analyze the role played by the significant overlap even when $T \to \infty$. The $K/T \to 0$ condition is quite general. The only case of interest that is not included is when K grows at the same rate as T, which shall be considered in the next section.

We state the large-sample properties of long-horizon predictive regressions under the $K/T \to 0$ approach in the following Theorem.

Theorem 3.1 *Under Assumptions 1-2,* $T \to \infty$, $K/T \to 0$, *and the null of no predictability, then*

(i) $\frac{\sqrt{T}}{K}\left(\hat{\beta}_K - \beta_K\right) \Rightarrow \frac{1}{\sqrt{c}}\left(\int_0^1 \tilde{J}_c(r)^2 dr\right)^{-1}\left(\int_0^1 \tilde{J}_c(r)\, dW_u(r)\right)$.

(ii) $\hat{\sigma}_K^2/K \to_p \sigma_u^2$.

(iii) $t/\sqrt{K} \Rightarrow \gamma\left(\int_0^1 \tilde{J}_c^2(r)\, dr\right)^{-1/2}\int_0^1 \tilde{J}_c(r)\, d\tilde{J}_c(r) + \left(1-\gamma^2\right)^{1/2}\mathcal{Z}$, $\mathcal{Z} \equiv \mathcal{N}(0,1)$, *independent of* $\tilde{J}_c(r)$.

(iv) $\frac{T}{K}\mathcal{R}_K^2 \Rightarrow F_1(W_u(r), J_c(r))$.

where $F_1(\cdot)$ *is a known functional of Ornstein-Uhlenbeck processes and nuisance parameters (see Appendix for details).*

Proof. *See Appendix* ■.

In Theorem 3.1, the convergence rate of the OLS estimator of β_K is a function of the overlap K. To understand the theorem, suppose that $K \sim T^{\alpha}$, $\alpha \in (0,1)$. The case of short-horizon regressions ($K = 1$) corresponds to $\alpha \to 0$. In this case, $T^{1/2}\left(\hat{\beta}_K - \beta_K\right) = O_p(1)$ and it follows that the estimator is consistent and converges at the usual rate to a well-defined distribution. For $\alpha \in (0, 0.5)$, the estimator retains consistency, but converges to a distribution at a rate slower than $T^{1/2}$. Consequently, a large number of observations would be necessary to reduce the sample bias. Finally, for $\alpha \geq 0.5$, the estimator is not-consistent and diverges with the sample size. The practical implication for a finite sample of size T is that, if K is larger than $T^{1/2}$, then $\hat{\beta}_K$ will not be an appropriate estimator of β_K. Moreover, in small samples the estimates $\hat{\beta}_K$ will tend to increase with K, even under the null of no predictability.

More importantly, the OLS t-statistic in long-horizon regressions diverges with the horizon K when the dependent variable is overlapped. In other words, as K increases we expect to see larger t-statistics even under the null of no predictability. This result is reminiscent of a spurious regression. It is due to the significant overlap and to the fact that the predictive variable is persistent. We conclude that the OLS t-statistic is not a suitable testing procedure when the overlap is a significant fraction of the sample size and when the predictor follows Assumptions (1) and (2). Note that inference in long-horizon regression is sometimes conducted with Newey-West standard errors by setting the truncation lag to equal $K-1$. We do not focus on this procedure because the characteristics of long-horizon regression, as well as the theoretical results obtained above, do not allow us to claim formally that the desired properties of this procedure holds in this context. The regularity conditions which insure consistency of the estimator of the long-run covariance matrix are not satisfied here. In particular, consistency follows if the truncation lag, say l_T, diverges with T and if $l_T/T \to 0$, as $T \to \infty$. The latter condition is generally not satisfied when $l_T = K - 1$. Furthermore, it is required that $T^{1/2}\left(\hat{\beta}_K - \beta_K\right) = O_p(1)$, which is not fulfilled in Theorem 3.1 above.[7]

Part (*iii*) of Theorem 3.1 proves that the re-scaled $t/\sqrt{K}$-statistic converges to a well-defined distribution. Hence, in regressions with significant overlaps, the $t/\sqrt{K}$-statistic should be used to test the null of no predictability. From (*iii*), $t/\sqrt{K}$ converges to a mixture of distributions

[7]See conditions *iii*) and *v*) in Newey and West (1987; Theorem 2, pp. 705)

if $\gamma \neq 0$, and to the standard normal distribution otherwise. In the former case, both the non-centrality parameter c and the correlation γ have an effect on the distribution. When $\alpha \to 0$, the results in Cavanagh *et al.* (1995) obtain. The nuisance parameters need to be estimated in order to conduct inference. Stock (1991), Phillips, Moon, and Xiao (2001), and Valkanov (2003) discuss alternative methods to estimate c, whereas γ can be estimated from the sample correlation. When $\gamma = 0$, the limiting distribution is completely free of nuisance parameters and the relevant percentiles are well-known. Finally, if there is not predictability in the return series, the $\mathcal{R}_K^2$ of the regression converges in probability to zero, but its rate of convergence depends on K. Therefore, in small samples, small variability of x_t may result in large $\mathcal{R}_K^2$ even under the null hypothesis of no predictability.

3.2 Asymptotic Properties: K/T→ κ

Under the parameterization $K = [\kappa T]$, $0 < \kappa < 1$, the overlap is a fixed fraction of the sample size. Lanne (2002), Richardon and Stock (1998), and Valkanov (2003) use this alternative approach to model the overlap in long-horizon regressions. When $K/T \to \kappa$, K diverges at the same rate as T, and the relative degree of overlap remains significant, even asymptotically. The properties of long-horizon regressions when $K/T \to \kappa$ are stated in the following Theorem.

Theorem 3.2 *Under Assumptions 1-2, $T \to \infty$, $K/T \to \kappa$, and the null of no predictability, then*

(i) $\left(\hat{\beta}_K - \beta_K\right)/\sqrt{K} \Rightarrow \frac{\sqrt{\kappa}}{\sqrt{c}} \left(\int_0^{1-\kappa} \tilde{J}_c(r)^2 \, dr\right)^{-1} \int_0^{1-\kappa} \tilde{J}_c(r)\, \psi(\kappa, r)\, dr.$

(ii) $\hat{\sigma}_K^2/K = O_p(1)$.

(iii) $t/\sqrt{K} \Rightarrow G(W_u(r), \kappa, J_c(r))$.

(iv) $\mathcal{R}_K^2 \Rightarrow F_2(W_u(r), \kappa, J_c(r))$.

where $\psi(\kappa, r) = [W_u(\kappa + r) - W_u(r)]$ and $G(\cdot)$ and $F_2(\cdot)$ are known functionals of Ornstein-Uhlenbeck processes and nuisance parameters (see Appendix for details).

Proof. *See Appendix* ■.

From part (i) in Theorem 3.2, it follows that $\left(\hat{\beta}_K - \beta_K\right)/\sqrt{K} = O_p(1)$ and hence the OLS estimator is not consistent for β_K. This result differs from Valkanov (2003), who shows with the same asymptotic approximation that $\left(\hat{\beta}_K - \beta_K\right)$ is $O_p(1)$. The difference in the results is due to the small variability of the predictor. Indeed, Valkanov (2003) treats σ_ε^2 as fixed rather than as local-to-zero as we do in Assumption (2). Part (i) of Theorem 3.2 implies that the larger is the overlap, the larger is $\hat{\beta}_K$ likely to be. This property matches very well with the empirical findings that $\hat{\beta}_K$ increases monotonically with the horizon (e.g., see Campbell *et al.* 1997), which is something the result in Valkanov (2003) has trouble explaining.

The limiting distribution of the normalized estimation bias in part (i) of Theorem 3.2 is a non-standard distribution that depends on the nuisance parameters (κ, c, γ). Similarly to the $K/T \rightarrow 0$ asymptotic approach, even though the slope cannot be consistently estimated, the estimation bias converges to a well-defined distribution as long as it is properly scaled. In this case, the suitable normalization is K. As in the previous section, the OLS estimator of the residuals variance diverges and must be normalized by K for convergence. However, when $K/T \rightarrow \kappa$ the normalized estimator of the variance converges weakly to a distribution rather than to a fixed parameter.

Interestingly, the OLS t-statistic does not converge to a well defined distribution, but diverges at a rate $\sqrt{K}$ (which is equivalent to $\sqrt{T}$). In other words, we expect the t-statistic to increase with the overlap, even in the absence of a relation between returns and the predictor variable. The properly normalized $t/\sqrt{K}$-statistic converges weakly to a non-standard distribution with a complex representation. This result agrees with the $K/T \rightarrow 0$ approach in the sense that the limiting distributions are well-defined. Therefore, the normalized statistic $t/\sqrt{K}$ emerges as a unified, natural way of conducting inference in long-horizon regressions. The limiting distributions differ depending on the speed of divergence of the overlap K. How different they are will be analyzed in the Monte Carlo section.

Finally, the $\mathcal{R}_K^2$ in regression (7) does not converge in probability to a fixed quantity, but is a positive random variable. This measure completely loses its natural meaning in this context and cannot be used as an indicator of the goodness of fit of the model when there is a significant overlap in the data. It must be noted that the t-statistic and the $\mathcal{R}_K^2$, are scale-invariant. Moreover, the derived results do not hinge on the assumption that the signal from x_t is small. Indeed, similar calculations carry through without that assumption (Valkanov 2003). But since the estimated bias is affected by the small variability of x_t, this feature is expected to affect the power function of the t-statistic in small samples. This issue will be analyzed in the Monte Carlo section.

3.3 How to Conduct Inference in Practice?

Theorems 3.1 and 3.2 emphasize the strong impact of the overlap on the statistical properties of long-horizon regressions. Looking at the results in the previous sections, several natural questions emerge. First, are these approximations more useful than the traditional fixed-K normal asymptotics in testing for predictability in long-horizon regressions? If they are, which of the two approximations should we use? Finally, from a practical perspective, how can we deal with the nuisance parameters? Before answering these questions, it must be noted that the use of asymptotic theory must be measured by whether it approximates the small-sample properties of the statistics of interest and whether in helps to characterize the distributions in addition to Monte Carlo experiments.

3.3.1 Which Asymptotic Approximation to Use?

The $K/T \to 0$ and $K/T \to \kappa$ asymptotics are more appropriate at modelling the properties of long-horizon regressions, whereas the fixed asymptotics are more suitable to account for the properties of statistics where the overlap is only a modest part of the sample size. This point was first made by Richardson and Stock (1988) in the case of the variance ratio statistic. Torous *et al.* (2004) make a similar observation when they compare the asymptotics of long-horizon regressions under the fixed-K and $K/T \to \kappa$ asymptotics. They conclude that the latter asymptotics provide a better approximation of the small-sample distributions when the overlap is a large fraction of the sample. This is not surprising given that the $K/T \to \kappa$ calculations emphasize the asymptotic importance of the overlap.

The question of whether to use the $K/T \to 0$ or $K/T \to \kappa$ asymptotics is difficult to answer from solely looking at Theorems 3.1 and 3.2, because the properties of the OLS estimator $\hat{\beta}_K$, its t-statistic, and the $\mathcal{R}_K^2$ are quite similar under both approximations. For instance, both Theorems imply that the values of $\hat{\beta}_K$ will tend to increase with the horizon under the null of no predictability. The t-statistic will also increase with the overlap under the two approximations even under the null hypothesis. Finally, the rate of convergence of $\mathcal{R}_K^2$ under the two asymptotics will depend on the overlap. Hence, both approximations imply that inference based on standard methods is not reliable when the overlap is a large fraction of the sample size. Interestingly, the $K/T \to 0$ or $K/T \to \kappa$ approximations show that a rescaled $t/\sqrt{K}$ statistic converges to a well-defined distribution. However, the limiting distributions under the two asymptotic calculations are different. Which distribution is better at approximating the small sample properties of the $t/\sqrt{K}$ statistic is ultimately an empirical question that we investigate with Monte Carlo simulations in the next section.

3.3.2 Nuisance Parameters

From an empirical perspective, the relevant question is how to carry out inference with the $t/\sqrt{K}$ statistics. In the theoretical section, we showed that the limiting distribution (and the critical values) of the $t/\sqrt{K}$ test depend on several parameters, such as the contemporaneous correlation γ, the degree of overlap as measured by either α or κ, and the value of the non-centrality parameter c. Therefore, given $(\gamma, \alpha, \kappa, c)$, the critical values can readily be obtained through standard Monte Carlo simulations. While (α, κ) are predetermined exogenously and γ can be inferred straightforwardly from the data, the parameter c pauses some challenges.[8]

There have been a few approaches of dealing with the c parameter. For instance, Stock (1991) proposes a median-unbiased estimator of c and centered confidence intervals which are

[8]Note that the OLS estimator of c is given by $\hat{c} = T(1 - \hat{\rho})$, where $\hat{\rho}$ is the estimation of the autoregressive parameter. From this, it is easy to see that $\hat{c}$ cannot be estimated consistently in this fashion. Rather, it converges weakly to a distribution rather than to the true c. For more details, see Phillips, Moon, Xiao (2001), Valkanov (2003) and references therein.

obtained by inverting the distribution of the (augmented) Dickey-Fuller statistic computed on x_t. This method is easy to implement, but it is unfortunately known to generate excessively large confident intervals. More importantly, the estimates of c can take non-negative values, which suggests that the predictor is an integrated or even an explosive process. Cavanagh *et al.* (1995) propose an alternative method that essentially circumvents the point-estimation of the nuisance parameter c. Their strategy is based on computing the test statistic for a wide range of values of c, and then extract conclusions on the basis of the supremum and infimum values of the test in this range (so-called sup-bounds test). The main drawback of this procedure, nevertheless, is that the resulting confident intervals prove to be too conservative and the resulting inference exhibits low power. The sup-bounds test is biased to accept the null hypothesis too frequently.

We focus on the following two methods of estimating c based on recent work by Valkanov (2003) and Phillips, Moon and Xiao (2001). Valkanov (2003) imposes a long-horizon restriction implied by an economic model in order to consistently estimate the parameter c. In the context of long-horizon regression, the long-horizon restriction is obtained from the Campbell and Shiller's (1988) log-linear model . The idea behind this procedure is similar in spirit to other approaches that incorporate prior modelling restrictions in order to improve the statistical properties of an estimator. For instance, the shrinkage approach in Ledoit and Wolf (2002) combines a standard sample estimator and an estimator implied from Sharpe's single-index market model to yield more accurate estimates of the variance of a portfolio.[9] More generally, the estimates implied from asset pricing models are widely used to solve the portfolio selection problem. In the specific context of long-horizon regressions, Valkanov (2003) exploits the information conveyed by the log-linearized Campbell and Shiller (1988) model to estimate c consistently. He then obtains the proper critical values for the $t/\sqrt{K}$-statistic. We shall provide more details

[9] The shrinkage procedure is well-known in decision theory. It is designed to balance the properties of two alternative estimators. In Ledoit and Wolf (2002), these are the sample covariance matrix, and the covariance matrix implied from the Sharpe's market model. Instead of choosing one of them, the procedure takes a convex linear weighted average of the two estimators. The optimal weights are determined according to some optimality criterion.

about this procedure below when we address the predictability of REIT returns.

Phillips, Moon and Xiao (2001) propose a cross-section procedure to estimate c consistently. Their idea is to divide the original sample into M non-overlapping time intervals which are then used to construct a panel data. The estimate of c is obtained by exploiting both the cross-sectional and time-series properties of the data. A disadvantage of this procedure is that the obtained estimate of c converges very slowly to the true parameter value. In small samples, this is reflected in imprecise estimates. We use this approach as a robustness check to the Valkanov (2003) method because it also has one advantage, namely, it it not based on a structural model. More importantly, in sections 4 and 5, we show that the critical values of our results are not particularly sensitive to c.

4 Small-Sample Properties: Monte Carlo Analysis

In this section we turn to the small-sample properties of the $t/\sqrt{K}$-statistic analyzed theoretically in the previous section. Our main goal is to explore whether the asymptotic results accurately characterize its small-sample properties in long-horizon regressions. In the Monte Carlo simulations, one-period returns (r_t) and dividend yields (x_t) are generated following equations (1)-(4). Without loss of generality, we can assume that the innovation terms $\eta = (u_t, \varepsilon_t)'$ are generated from a bivariate normal distribution with zero mean and covariance matrix $\Omega = \{\omega_{ij}\}$, and set $\omega_{11} = 1$ and $\omega_{22} = \omega_{11}c/T$. The covariance elements are given by $\omega_{ij} = \gamma\sqrt{c/T}$, where $\gamma = \{0, -0.90\}$, $c = \{0.5, 2, 5, 10, 20\}$ and $T = \{200, 500, 1000\}$. We do not consider samples smaller than 200 observations because the overwhelming majority of the studies on predictability use at least as many observations. A value of $\gamma = 0$ corresponds to an independent sequence, while $\gamma = -0.90$ allows for a strong degree of negative correlation between the innovations to the dividend yield and returns (Stambaugh 1999). The grid of c is chosen to correspond to realistic persistence and variances of the dividend yield, given the considered sample sizes. We conduct the simulations under the assumption that the nuisance parameters are known, because the primary goal of this exercise is to illustrate the small sample properties of the statistics of interest. Using the above data generating process, we define three Monte Carlo experiments.

Experiment I

The first experiment explores the speed of convergence of the $t/\sqrt{K}$ statistic for different degrees of overlapping as the sample size grows. The overlap is defined as either $K = [T^{\delta}]$ or $K = [\delta T]$ for $\delta = \{0.25, 0.50, 0.75\}$ and the three sample lengths specified above. In this experiment, we switch notation from α and κ to δ to emphasize that both of these parameters take the same values, and for a given sample size they imply different horizons K. The K-period returns are computed as $R_{K,t+1} = \sum_{j=0}^{K-1} r_{t+1+j}$. The long-horizon regressions are then estimated 25,000 times for all configurations $\{\delta, T, c, \gamma\}$. Since the estimated $\hat{\beta}_K$ in empirical tests are always positive and inference is conducted on the upper tail, we focus on the 95 percentile of the empirical distributions.[10]

[Insert Table 1 about here]

Tables 1 presents the 95 percentile under the $K = [T^{\delta}]$ approximation and Table 2 contains the same results under $K = [\delta T]$. In Table 1, we observe a fast convergence from the small sample of 200 observations to a large sample of 1000 observation. For different values of c, γ, and δ, the 95 percent quantiles of the $t/\sqrt{K}$-statistic are very similar as the sample size increases. Moreover, as proven in the asymptotic results, when $\gamma = 0$ the distribution is the standard normal. For non-zero γ and c, the distribution is leptokurtic and the skewness depends on the values of the pair (c, γ). A negative correlation skews the distribution to the right.[11] Since the asymptotic distribution relies on $K/T \to 0$, the approximation is particularly good in moderately-sized samples when K is not very large. Interestingly, the proximity of c to zero (ρ close to one) influences the speed of convergence in small samples. The closer is c to zero, the faster is the convergence, especially for large values of δ.

[10] We concentrate on the 95th percentile for clarity of exposition. Other percentiles yield very similar conclusions. All the results from these simulations are available upon request.

[11] This result is due to the fact that in the one-period return model the OLS estimation bias of β is proportional to both the covariance of the noise terms and to the bias of the OLS estimator of ρ. A negative correlation in the error terms induces a positive bias in the slope. The basic feature of the short-horizon regression is reproduced in the long-horizon regression.

[Insert Table 2 about here]

In Table 2, we observe that for the $K = [\delta T]$ approximation, the distribution of $t/\sqrt{K}$ converges rapidly to the asymptotic limit. The differences in the 95 percentiles with 200 and $1,000$ observations are very small, regardless the fraction κ. We notice that different values of c and δ have a more sizeable influence on the distributions than in the $K/T \to 0$ case. Interestingly, if δ is small (e.g. 0.25) and $\gamma = 0$, the limiting distribution as $c \to 0$ seems to be close to the standard normal distribution. This result is difficult to see from the asymptotic results. However, departures from normality arise fairly quickly as the autoregressive root in x_t decrease (c increases). As δ increases, the variance of the distribution decreases and the kurtosis is higher. Similarly to Table 1, contemporaneous correlation in the shocks skews the distributions considerably. The results from Tables 1 and 2 suggest that the $K = \left[T^{\delta}\right]$ and $K = [\delta T]$ approximations are both useful even in samples as small as 200 observations.

Experiment II

The purpose of this experiment is twofold. First, we analyze whether the relevant critical values implied by the distributions of $t/\sqrt{K}$ under the two approximations are different. Second, we assess whether the frequencies of rejection (empirical sizes) of $t/\sqrt{K}$ in a small sample correspond to the nominal ones when the critical values are generated from approximating the asymptotic distribution using a large number of observations, say T^*, where $T^* = 5000$. This experiment has clearly a practical orientation and we set the "driving parameter" K, to values of empirical relevance. We focus on $K = \{12, 24, 48, 60\}$ which correspond to return horizons of 1 to 5 years from a monthly dataset.

In Table 3, we present the asymptotic critical values of $t/\sqrt{K}$ from the distributions in Theorems 3.1 and 3.2. For the theory $K = [\kappa T]$, the critical values are obtained for various values of (κ, c, γ). The $K = [T^{\alpha}]$ approximations are simulated for $\alpha = 0.1$ and several value of (c, γ). Recall that in Theorem 3.1 we showed that the asymptotic distribution does not depend on the particular value of α and larger values of this parameter yield similar results. First, it is immediately clear that for a given value of K the asymptotic values under both approximations are very similar. In fact, for $\gamma = 0$ and $c \to 0$, the distributions are close to the standard normal for moderate degrees of overlap. But even for non-zero values of γ and c, the critical values under the two approximations are not very different. These similarities in the distributions suggests that the standardization of the t-statistic by $\sqrt{K}$ yields a statistic which is suitable for empirical tests. Once the statistic is normalized, the decision of whether to use the $K/T \to 0$ or the $K/T \to \kappa$ theory seems of secondary order of importance.

[Insert Tables 3, 4, and 5 about here]

Tables 4 and 5 display the empirical sizes of $t/\sqrt{K}$ using a sample of 500 observations. The asymptotic critical values are from Table 3. Since $T = 500$, the overlaps $K = \{12, 24, 48, 60\}$ lead to the same ratio K/T for both set of theories, thus implying different values of α and κ. The data are generated as explained above, the corresponding long-horizon regressions are estimated $25,000$ times for any configuration, and the fraction of $t/\sqrt{K}$-statistics that are beyond the corresponding critical values is reported in the tables. We observe that the empirical sizes under both approximations are very close to the nominal sizes. We notice increasing size distortion in Table 4 as c increases. Similar distortions are observed for increasing correlation γ. Importantly, the size distortions in Table 5 are less sensitive to the c and γ parameters.

The message that emerges from these simulations is that both approximations adequately capture the small sample distribution of $t/\sqrt{K}$. While there are some size distortions in the $K = [T^{\alpha}]$ approximation, they are not significant in comparison with the distortions that obtain when the t-statistic is not standardized by K. The $K = [\kappa T]$ theory seems to provide more robust approximation to the small sample distribution especially for large values of c and γ and at long horizons.

Experiment III

In the last experiment, we analyze the power properties of the $t/\sqrt{K}$ test under the same set of conditions as described in Experiment II. When the true process is generated under some non-zero β, we tabulate the probability of rejecting the null using the previously simulated critical values. We use the power simulations to address two issues.

Firstly, we analyze the impact of c on the power function. Recall that c enters into the persistence of the predictor and also in its volatility. The smaller is c, the more persistent and the less volatile is x_t. The effect of c on the power are displayed in Figures 1 ($K/T \to 0$) and 2 ($K/T \to \kappa$). For clarity of exposition and conciseness, we present only the results for $\gamma = -0.90$ and the sequence $c = [1/2, 5, 20]$ for the 95% confidence level. We focus on the predictability in one- and five-years horizons ($K = 12, 60$) using samples of 500 observations. Other values of c, γ, and T yield similar results. The estimated power functions in Figures 1 and 2 show that under both asymptotic approximations the test is consistent about identifying the alternative for different values of β. The higher the true value of β, the more likely it is to reject the null of non-predictability. Whether we are looking at a 12-period or 60-period horizon, a smaller c results in lower power. In other words, more persistent and more slowly varying processes will have a lower probability of rejecting the null when it is false. Interestingly, the drop in power is considerable when c is close to zero. Campbell (2001) and Campbell and Yogo (2003) reach a similar conclusion using different arguments.

[Insert Figures 1 and 2 about here]

Secondly, we consider the effect of the horizon on the power function. This simulation is of interest because it is often argued that the power of a test increases with the horizon. In Figures 3 and 4, we plot the power functions of the $t/\sqrt{K}$-statistic under the $K/T \to 0$ and $K/T \to \kappa$ approximations, respectively. We focus again on the results for $\gamma = -0.90$ and take the extreme values of $c = \{1/2, 20\}$ from our simulations. Under both theories, the size-adjusted shape of the power function is similar. When the signal-to-noise ratio is very low ($c = 1/2$), there is no noticeable increase in the power function as the horizon K increases. If we allow a stronger signal ($c = 20$) through a greater variability in the predictor, the power of the test increases under both theories. In fact, we observe a small "mechanical" drop in the power as the forecasting horizon increases, because the longer horizon implies fewer observations ($T - K$) available in the test, which ultimately leads to a reduction in power. If we disregard the loss of observations, an increase in the horizon does not lead to any relevant improvement in the power of the $t/\sqrt{K}$ test.

[Insert Figures 3 and 4 about here]

It is also interesting to compare our power results in Figure 4 to those in Valkanov (2003) who uses the same $K = [\kappa T]$ theory, but does not model the small variability of the predictor (equation 4). For similar values of c, γ, κ, and T, we obtain lower power functions, because the signal in our predictor is restricted to be weak. Not taking that restriction into account leads to seemingly higher power functions. We conclude that testing for predictability using long-horizon overlapping observations may be even more difficult than suggested in the previous literature.

5 Predictability of the REIT Portfolios

We use the results in the previous sections to test whether dividend yields paid out by Real Estate Investment Trusts (REITs) forecast future REIT returns. Three reasons prompted us to focus on REITs. First, the REIT market is rather new and unexplored compared to other sectors. The data on REITs goes back to only 1972, when the first REITs started trading on the open exchanges. More specifically, we have monthly data for the period 1972 to 2003 (384 observations) from NAREIT. Moreover, there were some important changes in the structuring of REITs in the early 1990s (Clayton and MacKinnon 2003; Glascock and Ghosh 2000). If we want to investigate long-horizon predictability with such small datasets, we want the small sample properties of our statistics to be accurately captured. This is something our approximations can deliver, as shown above. Second, REITs are required by law to pay out a large fraction (currently, at least a 90 percent) of their taxable income in the form of dividends. These features make REITs particularly suitable for testing time variation in expected returns. Predictability, if any, should be easier to detect in these assets as changes in the payout ratio are more likely to contain information about future returns. Third, the interest in REITs is likely to grow as these assets become more widely traded. While there are several studies on REITs (among others, Karolyi and Sanders 1998; Ling, Naranjo and Ryngaert 2000; Liu and Mei 1992; Nelling and Gyourko 1998) a consensus on their predictability is yet to emerge.

The entire NAREIT index includes currently about 180 publicly traded companies which can be divided into three portfolios: Equity-REITs (firms that invest in actual properties or other assets), Mortgage-REITs (corporations that loan money to real estate owners or invest in mortgage-backed securities), and Hybrid-REITs (firms that both own properties and make loans to real estate owners). We denote the entire index, Equity-REITs, Mortgage-REITs, and Hybrid-REITs by AREIT, EREIT, MREIT, and HREIT, respectively. Table 6 provides some descriptive statistics for both the monthly log returns and dividend yields of these REIT portfolios over the period 1972-2003, and the sub-periods 1972-1989 and 1990-2003. The REIT sector has experienced a dramatic growth and some transformations since the early 1990s, with deep changes affecting the behavior of returns. We therefore split the sample into these sub-periods to take into account possible structural breaks.[12]

[Insert Table 6 about here]

Since dividends are known to be highly seasonal, we filter them through a 12-month moving average, as in Fama and French (1988). The point estimates of the first-order autocorrelation parameter ρ yield values between 0.994 and 0.998 in all cases. The augmented Dickey-Fuller test is unable to reject the null of a unit root, but it is well-known that unit-root tests have low power against local-to-unity alternatives in finite samples. Returns series are serially uncorrelated in most cases, although some positive first-order correlation is evident for EREITs, and for two portfolios in the sub-period 1990-2003. While the dividend yields time-series are very persistent over the sample period, their annualized volatility is much smaller than the volatility of returns. Loosely speaking, their annualized standard deviation during the whole period is around 20-100

[12] Glascock and Ghosh (2000) report that, while only 13 firms existed prior to 1972, from 1990 to 1997 over 114 new REITS were created. Values increased from about \$6 billion in 1990 to over \$300 billion by the end of 1998. In addition, the sector has experimented several regulatory changes during the 1990s. Glascock, Lu and So (2000) analyze the effect of the 1993 Tax Act (effective on January 1, 1994), which favored institutional investment in the sector, reporting evidence of structural changes in the behaviour of returns. The REIT Modernization Act of 1999 (effective since January 1, 2001) stipulated, among others, a reduction in the mandatory payout requirement and changes in other regulatory requirements. Howe and Jain (2004) find structural changes in the pattern of the systematic risk of REITs.

times smaller than that of returns. The persistence and small variability of the dividend yield fits squarely with our Assumptions (1) and (2).

It is interesting to note in Table 6 that the dividend yield of Mortgage-REITs is significantly more volatile than the AREIT and the other portfolios. Therefore, if some of that variability contains information about future returns, we expect *a priori* MREITs to be more predictable than the entire index and the other REITs, everything else being equal.

[Insert Table 7 about here]

Table 7 presents the predictability results for horizons of 1 month to 5 years ($K = 1, 6, 12, 36, 60$) using the log of the dividend yield as a predictor. Predictability is addressed by using the entire sample and, as additional robustness check, two sub-periods.[13] The columns in the table display the estimates $\hat{\beta}_K$ of the slope coefficient, the t-statistic (computed with OLS and Newey-West errors [NW] with Bartlett kernel and truncation lag set to be equal to $\max\{1, K-1\}$), the normalized t-statistic $t/\sqrt{K}$ as suggested in the theoretical results, and the $\mathcal{R}_K^2$ of the regression. It is immediately clear that the results based on the standard regression statistics suggest strong predictability. As the return horizon increases, so do the estimated OLS coefficient β_K, the OLS and NW t-statistics, and the coefficient of determination $\mathcal{R}_K^2$. Similar results are reported in Campbell *et al.* (1997) and Fama and French (1988) for the value-weighted market portfolio.

The estimates $\hat{\beta}_K$ increase with the horizon K for most portfolios and periods. This pattern is typically observed in long-horizon regressions and some researchers present this finding as evidence that predictability is easier to detect at long horizons. However, as we showed in Theorems 3.1 and 3.2, larger observations of $\hat{\beta}_K$ are likely to be observed in regressions with significant overlap even under the null of no predictability. Hence, the large estimates of $\hat{\beta}_K$ at longer horizons do not necessarily imply that returns are predictable.

Similar observations apply to the OLS and NW t-statistics, which are increasing with the horizon for most portfolios and periods. If we rely on these two statistics and asymptotic

[13] The same qualitative pattern arises when we considered, for instance, two sub-periods of equal length. These results are not presented for saving space, but available from authors upon request.

normal critical values, the null of no predictability will easily be rejected. However, we proved in Theorems 3.1 and 3.2 that the OLS t-statistic will increase with the overlap even under the null of no predictability. Moreover, the NW t-statistic does not converge to a standard normal distribution (Torous *et al.* 2004) and may not be able to correct for the extreme dependence in the series in the context of long-horizon regressions. Also, since the $\mathcal{R}_K^2$ is a transformation of the OLS t-statistic, the same considerations apply. Hence, the increasing t-statistics and $\mathcal{R}_K^2$ cannot be taken as formal evidence for or against predictability in long-horizon regressions.

Interestingly, the $t/\sqrt{K}$-statistic does not exhibit nearly as much drift as K increases. This is expected as Theorems 3.1 and 3.2 both show that this statistic converges to a well-defined distribution. Moreover, in the Monte Carlo simulations we showed that the convergence was quite fast. The critical values of the $t/\sqrt{K}$-statistic are obtained as in the simulations under the approximation $K = [\kappa T]$, because of its overall small-sample properties. In simulating the critical values, we use a sample size $T^* = 5000$. From the invariance properties of the statistic, it is only necessary to set appropriate values for the nuisance terms (γ, κ, c). The parameter γ is estimated by its sample analog, and κ is obtained by the K/T ratio.

We estimate c by using two different methods. On the one hand, we estimate c by means of the Valkanov (2003) approach by imposing a long-horizon restriction from the Campbell and Shiller's (1988) model. Valkanov (2003) shows that a T-consistent estimate of c is given by $\hat{c} = \log\left(1 - \hat{\beta}_K\right)/\kappa$, $\kappa > 0$. We compute this estimator for each REIT portfolio using the entire sample and the estimate $\hat{\beta}_K$ for the longest horizon in our analysis. For the sake of robustness, we consider another consistent estimator of c by using the method in Phillips *et al.* (2001). This estimator is based on constructing a number of non-overlapping blocks which are then used to estimate c in a pooled fashion. In the implementation of this approach, we use 25 blocks of equal length on the entire sample for each portfolio.[14] The significance of the $t/\sqrt{K}$-statistic is denoted by $[A], \{\mathcal{A}\}$ (at the 1% level) $[B], \{\mathcal{B}\}$ (at the 5% level) and $[C], \{\mathcal{C}\}$ (at the 10% level) when the critical values are obtained with estimates of $\hat{c}$ using the Valkanov (2003) and Phillips *et al.* (2001) approaches, respectively.

[14] The Phillips *et al.* (2001) paper does not specify a method for determining the optimal number of blocks. We used different number of blocks, ranging from 20 to 30. The conclusions are qualitatively similar.

The overall evidence that emerges from the $t/\sqrt{K}$-statistic is that the dividend yield predicts future REIT returns. This is true in our sample and the sub-samples for most portfolios. However, the evidence for the predictability is not nearly as overwhelming as one would believe if only considering the standard t-statistics. In fact, the null of predictability is often rejected at the 5% level, but in many instances it cannot be rejected at the 1% level. Interestingly, for MREITs, the statistical evidence of predictability is stronger at short horizons. Ang and Bekaert (2003) have a similar finding for the returns of the aggregate stock market. In general, our results are consistent with previous findings about the existence of predictability in REIT returns.

Another important finding from the $t/\sqrt{K}$-statistics in Table 7 is that the predictability results at short- and long-horizons are not very different. In particular, we do not observe the increase in power that previous papers have reported. This is mainly due to the fact that we are explicitly modelling the small signal in the predictor in order to get the relevant critical values. Ignoring the small variability of x_t in the theorems and in the applications produces more powerful tests. As argued from theoretical and empirical perspectives (Table 6), this assumption seems crucial in providing correct inference in long-horizon predictability tests.

6 Conclusion

We propose a parsimonious way of capturing the persistence and small variability of conditioning variables, such as the dividend yield, that are often used in forecasting market returns. The predictor in forecasting regressions is modelled as having a highest autoregressive root close to unity and a small variance. We use this new process to derive the asymptotic properties of long-horizon predictive regressions with overlapping observations. In modelling the overlap K as a significant fraction of the sample size T, we consider asymptotic results under the assumptions $K/T \to 0$ and $K/T \to \kappa$, as $T \to \infty$. These two cases, corresponding to K increasing at a rate slower than and equal to the sample size, cover all reasonable asymptotic treatments of the overlap.

We derive the properties of long-horizon regressions under the $K/T \to 0$ and $K/T \to \kappa$

assumptions. We show that in both cases, the OLS estimator $\hat{\beta}_K$ of the predictive coefficient is not adequately estimated. Rather we are likely to observe larger estimates of $\hat{\beta}_K$ even under the null of no predictability. More importantly, we prove that the t-statistic also diverges under the null of no predictability, which can lead to spurious results and incorrect inference. Hence, the usual t-statistic cannot be used to test for predictability in long-horizon regressions with overlapping observations. Finally, we also derive the properties of the $\mathcal{R}^2_K$ in these predictability regressions.

A rescaled version of the t-statistic, $t/\sqrt{K}$, converges to a well defined distribution under the $K/T \to 0$ and $K/T \to \kappa$ asymptotics. Interestingly, Monte Carlo simulations suggest that the approximate distributions of the $t/\sqrt{K}$-statistic under the two approximations are quite similar for various reasonable values of the nuisance parameters. More importantly, both theories approximate well the small sample properties of the $t/\sqrt{K}$-statistic for large overlaps, with perhaps a slight edge for the $K/T \to \kappa$ calculations at very large values of K and for some values of the nuisance parameters. At long horizons, we do not detect an increase in power in the $t/\sqrt{K}$-statistic once the large overlap and the small variability of the predictor are taken into account.

We use the $t/\sqrt{K}$-statistic to test for long-horizon predictability in REITs. While we find that the dividend yield forecasts future REIT returns, the results are not nearly as strong as traditional inference approaches would suggest. The difference in the results is due to the failure of previous approaches to take into account the effect of the persistent and slowly-varying predictor, as well as the significant overlap on the small-sample properties of long-horizon regressions.

Appendix: Mathematical Proofs

To prove Theorem 3.1 we firstly state an useful Lemma.

Lemma 6.1 *Under Assumptions 1-2 the following results hold*

(i) $\sum_{t=1}^{T}\sum_{j=0}^{K-1} u_{t+1+j}x_t = K\sum_{t=1}^{T} x_t u_{t+1} + O_p\left(K\sqrt{K}\right) + O_p\left(\frac{K^2}{\sqrt{T}}\right).$

(ii) $\sum_{t=1}^{T}\sum_{j=0}^{K-1} u_{t+1+j} = K\sum_{t=1}^{T} u_{t+1} + O_p\left(K\sqrt{K}\right).$

(iii) $\frac{1}{TK}\sum_{t=1}^{T} U_{K,t+1}^2 \rightarrow_p \sigma_u^2.$

Proof

(i) Using the Beveridge-Nelson (BN) decomposition of $U_{K,t+1}$,

$$U_{K,t+1} = Ku_{t+1} + \sum_{j=0}^{K-2}(K-1-j)(u_{t+j+2} - u_{t+j+1}),$$

we write

$$\sum_{t=1}^{T}\sum_{j=0}^{K-1} u_{t+1+j}x_t = K\sum_{t=1}^{T} x_t u_{t+1} + \sum_{j=0}^{K-2}(K-1-j)\sum_{t=1}^{T} x_t(u_{t+j+2} - u_{t+j+1}). \tag{11}$$

Now we can rearrange the second term of this expression as,

$$\sum_{j=0}^{K-2}(K-1-j)\sum_{t=1}^{T} x_t(u_{t+j+2} - u_{t+j+1}) = I_{a1} - I_{a2} - I_{a3},$$

with

$$\begin{aligned} I_{a1} &= \sum_{j=0}^{K-2}(K-1-j)\, x_T u_{T+j+2},\ I_{a2} = \sum_{j=0}^{K-2}(K-1-j)\, x_1 u_{j+2}. \\ I_{a3} &= \sum_{j=0}^{K-2}(K-1-j)\sum_{t=1}^{T-1}(x_{t+1} - x_t)\, u_{t+j+2}. \end{aligned}$$

Note that,

$$E\left(I_{a1}^2\right) \leq \sigma_u^2 E\left(x_T^2\right)\sum_{j=0}^{K-2}(K-1-j)^2 = O\left(K^3\right),$$

so it follows from Chebyshev's inequality that $I_{a1} = O_p\left(K\sqrt{K}\right)$. Similarly, it can be shown that $I_{a2} = O_p\left(K\sqrt{K}\right)$.

From $x_{t+1} - x_t = -\frac{c}{T} x_t + \sigma_u \sqrt{\frac{c}{T}} e_{t+1}$ and Assumption 2 we have:

$$\begin{aligned} I_{a3} &= \sum_{j=0}^{K-2} (K-1-j) \sum_{t=1}^{T-1} (x_{t+1} - x_t) u_{t+j+2}, \\ &= -\frac{c}{T} \left[\sum_{t=1}^{T-1} x_t \left(\sum_{j=0}^{K-2} (K-1-j) u_{t+j+2} \right) \right] + \frac{\sigma_u \sqrt{c}}{\sqrt{T}} \sum_{t=1}^{T-1} e_{t+1} \left(\sum_{j=0}^{K-2} (K-1-j) u_{t+j+2} \right), \\ &= -I_{a31} + I_{a32}, \text{ say.} \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} I_{a31} &= \frac{c}{T} \sum_{j=0}^{K-2} (K-1-j) \left(\sum_{t=1}^{T-1} x_t u_{t+j+2} \right) \\ E(I_{a31}) &\leq \frac{c}{\sqrt{T}} \sqrt{\sum_{j=0}^{K-2} (K-1-j)^2} \sqrt{\sum_{j=0}^{K-2} E \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} x_t u_{t+j+2} \right)^2}. \end{aligned}$$

so, by noting $\sum_{j=0}^{K-2} (K-1-j)^2 = O(K^3)$ and $\sum_{j=0}^{K-2} E \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T-1} x_t u_{t+j+2} \right)^2 = O(K)$, it follows that $I_{a31} = O_p \left(\frac{K^2}{\sqrt{T}} \right)$. Since,

$$E\left(I_{a32}^2\right) = c\sigma_u^3 \sum_{j=0}^{K-2} (K-1-j)^2 = O\left(K^3\right),$$

we have $I_{a32} = O_p \left(K\sqrt{K} \right)$. Therefore, $I_{a3} = O_p \left(K\sqrt{K} \right) + O_p \left(\frac{K^2}{\sqrt{T}} \right)$.

Combining I_{a1}, I_{a2}, and I_{a3}, we have

$$K \sum_{j=0}^{K-2} \left(\frac{K-1-j}{K} \right) \sum_{t=1}^{T} x_t (u_{t+j+2} - u_{t+j+1}) = O_p \left(K\sqrt{K} \right) + O_p \left(\frac{K^2}{\sqrt{T}} \right),$$

and in consequence, in view of (11), we have the required result,

$$\sum_{t=1}^{T} \sum_{j=0}^{K-1} u_{t+1+j} x_t = K \sum_{t=1}^{T} x_t u_{t+1} + O_p \left(K\sqrt{K} \right) + O_p \left(\frac{K^2}{\sqrt{T}} \right).$$

■

(ii) Replacing $x_t = 1$ in (11) and proceeding as in I_{a1} and I_{s2}, we may have

$$\begin{aligned} \sum_{t=1}^{T} \sum_{j=0}^{K-1} u_{t+1+j} &= K \sum_{t=1}^{T} u_{t+1} + \sum_{j=0}^{K-2} (K-1-j) \sum_{t=1}^{T} (u_{t+j+2} - u_{t+j+1}) \\ &= K \sum_{t=1}^{T} u_{t+1} + \sum_{j=0}^{K-2} (K-1-j) (u_{T+j+2} - u_{j+2}) = K \sum_{t=1}^{T} u_{t+1} + O_p \left(K\sqrt{K} \right). \end{aligned}$$

as required. That completes the proof. ■

(iii) Apply the BN decomposition of Phillips and Solo (1992) to $U_{K,t+1}$ and write

$$\begin{aligned} U_{K,t+1}^2 &= Ku_{t+K}^2 + \left(\tilde{U}_{aK,t-1} - \tilde{U}_{aK,t}\right) \\ &\quad +2u_{t+K}\left(\sum_{j=1}^{K-1}(K-j)\,u_{t+K-j}\right) + 2\left(\tilde{U}_{bK,t-1} - \tilde{U}_{bK,t}\right), \end{aligned}$$

where

$$\begin{aligned} \tilde{U}_{aK,t} &= \sum_{j=0}^{K-2}(K-1-j)\,u_{t+K-j} \\ \tilde{U}_{bK,t} &= \sum_{j=1}^{K-2}\sum_{l=0}^{K-2-j}(K-1-j-l)\,u_{t+K-l}u_{t+K-j-l}. \end{aligned}$$

Using this, we express

$$\begin{aligned} \frac{1}{TK}\sum_{t=1}^{T}U_{K,t+1}^2 &= \frac{1}{T}\sum_{t=1}^{T}u_{t+K}^2 + 2\frac{1}{T}\sum_{t=1}^{T}\left(u_{t+K}\left(\sum_{j=1}^{K-1}\left(\frac{K-j}{K}\right)u_{t+K-j}\right)\right) \\ &\quad +\frac{1}{TK}\left(\tilde{U}_{aK,0} - \tilde{U}_{aK,T}\right) + 2\frac{1}{TK}\left(\tilde{U}_{bK,0} - \tilde{U}_{bK,T}\right) \\ &= I_c + 2II_c + III_c + 2IV_c, \text{ say.} \end{aligned}$$

Notice by the SLLN, we have

$$I_c \to_{a.s.} \sigma_u^2.$$

The required result follows if we show $II_c, III_c, IV_c \to_p 0$ when $K/T \to 0$. For II_c, notice that

$$E\left(II_c^2\right) \leq \frac{2}{T^2K^2}E\left(\tilde{U}_{aK,0}^2 + \tilde{U}_{aK,T}^2\right).$$

A direct calculation shows that

$$\begin{aligned} \frac{1}{T^2K^2}E\tilde{U}_{aK,0}^2 &= \frac{1}{T^2K^2}E\left(\sum_{j=0}^{K-2}(K-1-j)\,u_{t+K-j}\right)^2 = \sigma_u^2\frac{1}{T^2K^2}\sum_{j=0}^{K-2}(K-1-j)^2 \\ &= O\left(\frac{K}{T^2}\right) = o(1). \end{aligned}$$

So, $\tilde{U}_{aK,0}/TK = o_p(1)$. Similarly, we can show that $\tilde{U}_{aK,T}/TK = o_p(1)$, and we have the required result for II_c. For III_c, notice by the Cauchy-Schwarz inequality that

$$
\begin{aligned}
E\left(III_c\right)^2 &= \frac{K}{T}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{t+K}\left(\frac{1}{K}\sum_{j=1}^{K-1}\left(\frac{K-j}{K}\right)u_{t+K-j}\right)\right)^2 \\
&= \frac{K}{T}\left(\frac{1}{T}\sum_{t=1}^{T}E\left(u_{t+K}^2\right)E\left(\frac{1}{K}\sum_{j=1}^{K-1}\left(\frac{K-j}{K}\right)u_{t+K-j}\right)^2\right) \\
&= \frac{K}{T}O\left(1\right)=o\left(1\right),\ \text{as } \frac{K}{T}\to 0.
\end{aligned}
$$

Therefore,

$$
III_c = O_p\left(\sqrt{\frac{K}{T}}\right)=o_p\left(1\right),
$$

as required. For IV_c, we have

$$
\begin{aligned}
\frac{1}{T^2K^2}E\tilde{U}_{bK,0}^2 &= \frac{1}{T^2K^2}E\left(\sum_{j=1}^{K-2}\sum_{l=0}^{K-2-j}\left(K-1-j-l\right)u_{t+K-l}u_{t+K-j-l}\right)^2 \\
&= \frac{1}{T^2K^2}\sum_{j=1}^{K-2}\sum_{i=1}^{K-2}\sum_{l=0}^{K-2-j}\sum_{m=0}^{K-2-j}\left(K-1-j-l\right)\left(K-1-i-m\right) \\
&\quad\times E\left(u_{t+K-l}u_{t+K-m}u_{t+K-j-l}u_{t+K-i-m}\right) \\
&= \sigma_u^4\frac{1}{T^2K^2}\sum_{j=1}^{K-2}\sum_{l=0}^{K-2-j}\left(K-1-j-l\right)^2=O\left(\frac{K^2}{T^2}\right)=o\left(1\right).
\end{aligned}
$$

Similarly, we can show that $\frac{1}{T^2K^2}E\tilde{U}_{bK,T}^2=o\left(1\right)$. So, as for II_c, we can deduce that

$$
IV_c = o_p\left(1\right).
$$

as required, and we complete the proof. ■

Proof of Theorem 3.1

(i) Note that

$$\frac{\sqrt{T}}{K}\left(\hat{\beta}_K-\beta_K\right)=\frac{\frac{1}{K\sqrt{T}}\sum_{t=1}^{T-K}\left(\sum_{j=0}^{K-1}u_{t+1+j}\right)(x_t-\bar{x})}{\frac{1}{T}\sum_{t=1}^{T-K}(x_t-\bar{x})^2}=\frac{A_{T,K}}{B_T}.$$

From Lemma 3.1, the continuous mapping theorem, and the fact that $K/T\to 0$ it follows easily that

$$B_T\Rightarrow\sigma_u^2c\int_0^1\left(J_c(r)-\int_0^1J_c(r)\,dr\right)^2dr\overset{let}{=}\sigma_u^2c\int_0^1\tilde{J}_c^2(r)\,dr.$$

Next, from Lemma 6.1(i) and (ii) and by $K/T\to 0$, we have

$$A_{T,K}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T-K}(x_t-\bar{x})\,u_{t+1}+O_p\left(\sqrt{\frac{K}{T}}\right)+O_p\left(\frac{K}{\sqrt{T\,(T-K)}}\right)$$
$$=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\bar{x})\,u_{t+1}+o_p(1)\,.$$

Using similar arguments in the proof of Lemma 1(d) of Phillips (1987), we can derive $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(x_t-\bar{x})\,u_{t+1}\Rightarrow\sigma_u^2\sqrt{c}\int_0^1\tilde{J}_c(r)\,dW_u(r)$, from which we have

$$A_{T,K}\Rightarrow\sigma_u^2\sqrt{c}\int_0^1\tilde{J}_c(r)\,dW_u(r)\,.$$

Finally, from applying the continuous mapping theorem, we have the required result. ■

(ii) Denote

$$\xi_{T,K}=\sum_{t=1}^{T-K}\left[U_{K,t+1},x_tU_{K,t+1}\right]';\quad\Psi_{T,K}=\sum_{t=1}^{T-K}[1,x_t]'\,[1,x_t]\,.\tag{12}$$

Notice by definition and under no-predictability that,

$$\begin{aligned}\frac{1}{K}\hat{\sigma}_K^2&=\frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}\left(R_{K,t+1}-\hat{\alpha}_K-\hat{\beta}_Kx_t\right)^2\\&=\frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}U_{K,t+1}^2-\frac{1}{(T-K)\,K}\left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right).\end{aligned}$$

Denote $\tilde{\xi}_{T,K}=\frac{1}{K\sqrt{T-K}}\xi_{T,K}$ and $\tilde{\Psi}_{T,K}=\frac{1}{T-K}\Psi_{T,K}$ the corresponding renormalized terms. By Lemma 6.1(i) - (iii),

$$\frac{1}{(T-K)\,K}\left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right)=\frac{K}{T-K}\tilde{\xi}_{T,K}'\tilde{\Psi}_{T,K}^{-1}\tilde{\xi}_{T,K}=O_p\left(\frac{K}{T}\right),$$

and therefore

$$\frac{1}{(T-K)\,K}\sum_{t=1}^{T-K} U_{K,t+1}^2 \to_p \sigma_u^2.$$

Since $K/T \to 0$, we have the required result that

$$\frac{1}{K}\hat{\sigma}_K^2 \to_p \sigma_u^2.$$

That completes the proof. ■

(iii) Notice that

$$\frac{1}{\sqrt{K}}t = \left(\frac{\hat{\sigma}_K^2}{K}\right)^{-1/2}\left(\frac{\sum_{t=1}^{T-K}(x_t-\bar{x})^2}{T}\right)^{1/2}\frac{\sqrt{T}}{K}\hat{\beta}_K.$$

Then, under the null of $\beta_K = 0$, by Theorem 3.1(i) and (ii) and the continuos mapping theorem, it follows easily

$$\frac{t}{\sqrt{K}} \Rightarrow \frac{\int_0^1 \tilde{J}_c(r)\,dW_u(r)}{\left(\int_0^1 \tilde{J}_c^2(r)\,dr\right)^{1/2}}.$$

which can be rewritten alternatively as

$$\frac{t}{\sqrt{K}} \Rightarrow \gamma\left(\int_0^1 \tilde{J}_c^2(r)\,dr\right)^{-1/2}\int_0^1 \tilde{J}_c(r)\,dJ_c(r) + \left(1-\gamma^2\right)^{1/2}\mathcal{Z}.$$

That completes the proof.■

(iv) By definitions of $\mathcal{R}_K^2$,

$$\mathcal{R}_K^2 = 1 - \frac{\sum_{t=1}^{T-K} U_{K,t+1}^2 - \left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right)}{\sum_{t=1}^{T-K}\left(R_{K,t+1}-\bar{R}_K\right)^2},$$

where $\xi_{T,K}$ and $\Psi_{T,K}$ are defined in (12). Under the no-predictability assumption if follows that

$$\begin{aligned}\frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}\left(R_{K,t+1}-\bar{R}_K\right)^2 &= \frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}\left(U_{K,t+1}-\bar{U}_K\right)^2\\ &= \frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}U_{K,t+1}^2 - \frac{1}{K}\bar{U}_K^2\\ &= \frac{1}{(T-K)\,K}\sum_{t=1}^{T-K}U_{K,t+1}^2\left(1-m_{K,T}\right),\end{aligned}$$

where

$$
\begin{aligned}
m_{K,T} &= \left(\frac{1}{(T-K)K}\sum_{t=1}^{T-K}U_{K,t+1}^2\right)^{-1}\frac{1}{K}\bar{U}_K^2 \\
&= \frac{K}{(T-K)}\left(\frac{1}{(T-K)K}\sum_{t=1}^{T-K}U_{K,t+1}^2\right)^{-1}\left(\frac{1}{K\sqrt{T-K}}\sum_{t=1}^{T-K}U_{K,t+1}\right)^2.
\end{aligned}
$$

Denote

$$
l_{K,T} = \left(\frac{1}{(T-K)K}\sum_{t=1}^{T-K}U_{K,t+1}^2\right)^{-1}\frac{1}{(T-K)K}\left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right),
$$

Then,

$$
\mathcal{R}_K^2 = \left(1-\frac{1-l_{K,T}}{1-m_{K,T}}\right) = \frac{l_{K,T}-m_{K,T}}{1-m_{K,T}}. \tag{13}
$$

Notice that

$$
\begin{aligned}
\frac{T}{K}\left(l_{K,T}-h_{K,T}\right) &= \left(\frac{1}{(T-K)K}\sum_{t=1}^{T-K}U_{K,t+1}^2\right)^{-1} \\
&\quad\times\left(\frac{1}{K^2}\left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right)-\left(\frac{1}{K\sqrt{T-K}}\sum_{t=1}^{T-K}U_{K,t+1}\right)^2\right) \\
&\Rightarrow \xi'\left(\Psi^{-1}-E_{11}\right)\xi,
\end{aligned}
$$

where

$$
\begin{aligned}
\xi &= \left(\int_0^1 dW_u(r), \int_0^1 J_c(r)\,dW_u(r)\right)', \\
\Psi &= \begin{pmatrix} 1 & \int_0^1 J_c(r)\,dr \\ \int_0^1 J_c(r)\,dr & \int_0^1 J_c(r)^2\,dr \end{pmatrix},\ E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\end{aligned}
$$

Also, it is easy to find that

$$
m_{K,T} = O_p\left(\frac{K}{T}\right) = o_p(1).
$$

Therefore, in view of (13) and by Slutsky theorem, we have the required result.■

Proof of Theorem 3.2

(i) Note that

$$
\frac{1}{\sqrt{T}}\left(\hat{\beta}_K-\beta_K\right) = \frac{\frac{1}{T}\sum_{t=1}^{T-K}\left(\frac{1}{\sqrt{T}}\sum_{j=0}^{K-1}u_{t+1+j}\right)(x_t-\bar{x})}{\frac{1}{T}\sum_{t=1}^{T-K}(x_t-\bar{x})^2} = \frac{C_{T,K}}{D_T}.
$$

From the continuous mapping theorem, and similar to C_T, it follows easily from the fact $K/T \rightarrow \kappa$ that

$$D_T \Rightarrow c\sigma_u^2 \int_0^{1-\kappa} \tilde{J}_c^2(r)\, dr.$$

Observe that $C_{T,K}$ can be rewritten as

$$C_{T,K} = \frac{1}{T}\sum_{t=1}^{T-K}(x_t - \bar{x})\left[\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{[\kappa T]+[rT]} u_s\right) - \left(\frac{1}{\sqrt{T}}\sum_{s=1}^{[rT]} u_s\right)\right].$$

Denote $\psi(\kappa, r) = (W_u(\kappa + r) - W_u(r))$. From the continuous mapping theorem it follows that

$$C_{T,K} \Rightarrow \sqrt{c}\sigma_u^2 \int_0^{1-\kappa} \tilde{J}_c(r)\, \psi(\kappa, r)\, dr,$$

and hence

$$\frac{1}{\sqrt{T}}\left(\hat{\beta}_K - \beta_K\right) \Rightarrow \frac{1}{\sqrt{c}}\frac{\int_0^{1-\kappa} \tilde{J}_c(r)\, \psi(\kappa, r)\, dr}{\int_0^{1-\kappa} \tilde{J}_c^2(r)\, dr}.$$

and clearly$\left(\hat{\beta}_K - \beta_K\right) = O_p\left(\sqrt{T}\right)$ under the null of non-predictability. Equivalently,

$$\frac{1}{\sqrt{K}}\left(\hat{\beta}_K - \beta_K\right) \overset{d}{=} \sqrt{\kappa}\left[\frac{1}{\sqrt{T}}\left(\hat{\beta}_K - \beta_K\right)\right].$$

■

(ii) Consider the OLS estimator $\hat{\sigma}_K^2$ and note

$$\frac{1}{T}\hat{\sigma}_K^2 = \frac{1}{T(T-K)}\sum_{t=1}^{T-K}\left(R_{K,t+1} - \hat{\alpha}_K - \hat{\beta}_K x_t\right)^2 = \frac{1}{T(T-K)}\sum_{t=1}^{T-K}\hat{U}_{K,t+1}^2.$$

Now recall the definitions $\xi_{T,K} = \sum_{t=1}^{T-K}[U_{K,t+1}, x_t U_{K,t+1}]'$ and $\Psi_{T,K} = \sum_{t=1}^{T-K}[1, x_t]'[1, x_t]$ in (12). Under no-predictability we can rewrite

$$\frac{1}{T}\hat{\sigma}_K^2 = \frac{1}{T(T-K)}\sum_{t=1}^{T-K} U_{K,t+1}^2 - \frac{1}{T(T-K)}\left(\xi_{T,K}'\Psi_{T,K}^{-1}\xi_{T,K}\right).$$

From the continuous mapping theorem and similar to Theorem 3.1. it follows that

$$\frac{1}{T(T-K)}\sum_{t=1}^{T-K}\left(\sum_{j=0}^{K-1} u_{t+1+j}\right)^2 \Rightarrow \sigma_u^2\int_0^{1-\kappa}\psi^2(\kappa, r)\, dr \equiv \sigma_u^2\lambda_1,$$

also,

$$T^{-3/2}\xi_{T,K} \Rightarrow \left(\int_0^{1-\kappa} \sigma_u \psi^2(\kappa, r)\, dr, \int_0^{1-\kappa} \sigma_u^2 \sqrt{c} J_c(r) \psi(\kappa, r)\, dr \right)' \equiv \sigma_u \lambda_2',$$

and finally,

$$T^{-1}\Psi_{T,K} \Rightarrow \begin{pmatrix} 1 & \sigma_u \sqrt{c} \int_0^{1-\kappa} J_c(r)\, dr \\ \sigma_u \sqrt{c} \int_0^{1-\kappa} J_c(r)\, dr & \sigma_u^2 c \int_0^{1-\kappa} J_c^2(r)\, dr \end{pmatrix} \equiv \Psi.$$

Therefore,

$$\frac{\hat{\sigma}_K^2}{T} \Rightarrow \sigma_u^2 \left[\lambda_1 - \left(\lambda_2' \Psi^{-1} \lambda_2 \right) \right].$$

so clearly $\hat{\sigma}_K^2 = O_p(T)$. Alternatively,

$$\frac{\hat{\sigma}_K^2}{K} \overset{d}{=} \sqrt{\kappa} \left[\frac{\hat{\sigma}_K^2}{T} \right].$$

■

(iii) Consider the OLS t-statistic. From the above results, we can show that

$$t = \frac{\left(\sum_{t=1}^{T-K} (x_t - \bar{x})^2 \right)^{1/2} \left(\hat{\beta}_K - \beta_K \right)}{\hat{\sigma}_K} = O_p\left(\sqrt{T} \right).$$

and hence the statistic diverges with the sample size. Under a suitable normalizing,

$$\frac{1}{\sqrt{K}} t \Rightarrow \sqrt{\kappa} \left(\lambda_1 - \left(\lambda_2' \Psi^{-1} \lambda_2 \right) \right)^{-1/2} \left(\int_0^{1-\kappa} \tilde{J}_c(r) \psi(\kappa, r)\, dr \right) \left(\int_0^{1-\kappa} \tilde{J}_c^2(r)\, dr \right)^{-1/2}.$$

That completes the proof.■

(iv) From the definition of $\mathcal{R}_K^2$ and under the no-predictability assumption we can write,

$$\mathcal{R}_K^2 = 1 - \frac{(T(T-K))^{-1} \sum_{t=1}^{T-K} U_{K,t+1}^2 - (T(T-K))^{-1} \left(\xi_{T,K}' \Psi_{T,K}^{-1} \xi_{T,K} \right)}{(T(T-K))^{-1} \sum_{t=1}^{T-K} \left(U_{K,t+1} - \bar{U}_K \right)^2}.$$

for $\xi_{T,K}$ and $\Psi_{T,K}$ defined above. Clearly, the numerator of this expression will converge to the same limit distribution as $\hat{\sigma}_K^2/T$ in part (ii) of this Theorem. Finally, note that

$$\begin{aligned} \frac{1}{T(T-K)} \sum_{t=1}^{T-K} \left(U_{K,t+1} - \bar{U}_K \right)^2 &= \frac{1}{T(T-K)} \sum_{t=1}^{T-K} U_{K,t+1}^2 - \frac{1}{T} \bar{U}_K^2 \\ &\Rightarrow \sigma_u^2 \left[\int_0^{1-\kappa} \psi^2(\kappa, r)\, dr - \left(\int_0^{1-\kappa} \psi(\kappa, r)\, dr \right)^2 \right]. \end{aligned}$$

and hence $\mathcal{R}_K^2 = O_p(1)$ as $T \to \infty$. That completes the proof. ■

References

[1] Ang, A., and Bekaert, G. (2003), "Stock Return Predictability: Is It There?," NBER Working Paper.

[2] Baker, M., and Wurgler, J. (2000), "The Equity Share in New Issues and Aggregate Stock Returns," *Journal of Finance*, 55, 2219-2257.

[3] Baker, M., and Wurgler, J. (2004), "A Catering Theory of Dividends," *Journal of Finance*, 59, 1125-1165.

[4] Bossaerts, P., and Hillion, P. (1999), "Implementing Statistical Criteria to Select Return Forecasting Models: What Do We Learn?," *Review of Financial Studies*, 12, 405–428.

[5] Campbell, J.Y. (1987), "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

[6] − − − −(2001), "Why Long Horizons? A Study of Power Against Persistent Alternatives," *Journal of Empirical Finance*, 8, 459–491.

[7] Campbell, J.Y., Lo, A.W., and MacKinlay, A.C. (1997), *The Econometrics of Financial Markets*, Princeton NJ: Princeton University Press.

[8] Cambell, J.Y., and Shiller, R. (1988), "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 405-427.

[9] Campbell, J.Y., and Shiller, R. (1989), "The Dividend Ratio Model and Small Sample Bias: A Monte Carlo Study," *Economics Letters*, 29, 325-331.

[10] Campbell, J.Y., and Yogo, M. (2003), "Efficient Tests of Stock Return Predictability," NBER Working Paper.

[11] Cavanagh, C., Elliott, G., and Stock, J. (1995), "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11, 1131-1147.

[12] Clayton, J., and MacKinnon, G. (2003), "The Relative Importance of Stock, Bond and Real Estate Factors in Explaining REIT Returns," *Journal of Real Estate Finance and Economics*, 27, 39-60.

[13] Cochrane, J. H. (2001), *Asset Pricing*, Princeton NJ: Princeton University Press.

[14] Elliott, G. (1998), "The Robustness of Cointegration Methods when Regressors Almost Have Unit Roots," *Econometrica*, 66, 149-58.

[15] Fama, E.F., and French, K.R. (1988), "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3-25.

[16] Fama, E.F., and French, K. (2001), "Disappearing Dividends: Changing Firm Characteristics or Lower Propensity to Pay?," *Journal of Financial Economics*, 60, 3-43.

[17] Fama, E.F., and Schwert, G. (1977), "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115-146.

[18] Ferson, W.E., Sarkissian, S., and Simin, T.T. (2003), "Spurious Regression in Financial Economics?," *Journal of Finance*, 58, 1393-1413.

[19] Goetzmann, W.N., and Jorion, P. (1993), "Testing the Predictive Power of Dividend Yields," *Journal of Finance*, 48, 663-679.

[20] Glascock, J.L., and Ghosh, C. (2000), "The Maturation of a Developing Industry: REITS in the 1999s," *Journal of Real Estate Finance and Economics*, 20, 87-90.

[21] Glascock, J.L., Lu, C., and So R.W. (2000) "Further Evidence on the Integration of REIT, Bond, and Stock Returns," *Journal of Real Estate Finance and Economics*, 20, 177-194.

[22] Herrndorf, N. (1984), "A Functional Central Limit Theorem for Weakly Dependent Sequences of Random Variables," *Annals of Probability*, 12, 141-153.

[23] Hodrick, R. (1992), "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357-386.

[24] Howe, J.S., and Jain, R. (2004) "The REIT Modernization Act of 1999," *Journal of Real Estate Finance and Economics*, 28, 369-388.

[25] Karolyi, G.A., and Sanders, A.B. (1998), "The Variation of Economic Risk Premiums in Real Estate Returns," *Journal of Real Estate Finance and Economics*, 17, 245-262.

[26] Lanne, M. (2002), "Testing the Predictability of Stock Returns," *The Review of Economics and Statistics*, 84, 407-415.

[27] Ledoit, O., and Wolf, M. (2002), "Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection," *Journal of Empirical Finance,* 10, 603-621.

[28] Lettau, M., and Ludvigson, S. (2001), "Resurrecting the (C)CAPM: A Cross-Sectional Test when Risk Premia are Time-Varying," *Journal of Political Economy*, 109, 1238-1287.

[29] Lewellen, J. (2004) "Predicting Returns with Financial Ratios," *Journal of Financial Economics*, forthcoming.

[30] Ling, D.C., Naranjo, A., and Ryngaert, M.D. (2000), "The Predictability of Equity REIT Returns: Time Variation and Economic Significance," *Journal of Real Estate Finance and Economics*, 20, 117-136.

[31] Liu, C.H., and Mei, J. (1992), "The Predictability of Returns on Equity REITs and Their Co-Movement with Other Assets," *Journal of Real Estate Finance and Economics*, 5, 401-418.

[32] Mankiw, N., and Shapiro, M. (1986), "Do we reject too often? Small sample properties of tests of rational expectations models," *Economics Letters*, 20, 139-145.

[33] Nelson, C., and Kim, M. (1993), "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 641-661.

[34] Nelling, E., and Gyourko J. (1998), "The Predictability Of Equity REIT Returns," *Journal of Real Estate Research*, 16, 251-268.

[35] Newey, W., and West, K.D. (1987), "A Simple, Positive Semi-definite Heteroskedasticiy and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703-708.

[36] Phillips, P.C.B. (1987), "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

[37] Phillips, P.C.B., Moon, R., and Xiao, Z. (2001), "How to Estimate Autoregressive Unit Roots Near Unity", *Econometric Theory*, 17, 29-69.

[38] Phillips, P.C.B., and Solo, V. (1992) "Asymptotics for Linear Processes," *Annals of Statistics*, 20, 971-1001.

[39] Richardson, M., and Stock, J.H. (1989), "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics,* 25, 323-348.

[40] Shefrin, H.M., and Statman, M. (1984), "Explaining Investor Preference for Cash Dividends," *Journal of Financial Economics*, 13, 253-282.

[41] Shiller, R. J. (1981), "Do Stock Price Move too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71, 421-436.

[42] Stambaugh, R.F. (1986), "Bias in Regressions with Lagged Stochastic Regressors," University of Chicago Working Paper.

[43] ––––––(1999), "Predictive Regressions," *Journal of Financial Economics*, 54, 375-421.

[44] Stein, J.C. (1996), "Rational Capital Budgeting in an Irrational World," *Journal of Business*, 69, 429-455.

[45] Stock, J.H. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, 435-459.

[46] Summers, L. (1986), "Does the Stock Market Rationally Reflect Fundamental Values?," *Journal of Finance,* 41, 591-600.

[47] Torous, W., Valkanov, R., and Yan, S. (2004), "On Predicting Stock Returns with Nearly Integrated Explanatory Variables", *Journal of Business,* 77, 937-966.

[48] Torous, W., and Valkanov, R. (2003), "Boundaries of Predictability: Noisy Predictive Regressions," University of California, Los Angeles, Working Paper.

[49] Valkanov, R. (2003), "Long-Horizon Regressions: Theoretical Results and Applications," *Journal of Financial Economics*, 68, 201-232.

Table 1: Empirical Percentile $K = \left[T^{\delta}\right]$

$$\Pr\left(t/\sqrt{K} \leq \xi\right) = 0.95$$

Correlation $\gamma = 0$

	$\delta = 0.25$			$\delta = 0.50$			$\delta = 0.75$		
c	T=200	T=500	T=1000	T=200	T=500	T=1000	T=200	T=500	T=1000
1/2	1.651	1.646	1.649	1.672	1.669	1.664	1.615	1.666	1.686
2	1.635	1.632	1.669	1.640	1.637	1.671	1.533	1.591	1.624
5	1.602	1.642	1.667	1.623	1.636	1.638	1.375	1.474	1.506
10	1.625	1.653	1.629	1.581	1.582	1.597	1.175	1.264	1.330
20	1.559	1.595	1.629	1.362	1.454	1.514	0.910	1.004	1.083

Correlation $\gamma = -0.90$

	$\delta = 0.25$			$\delta = 0.50$			$\delta = 0.75$		
c	T=200	T=500	T=1000	T=200	T=500	T=1000	T=200	T=500	T=1000
1/2	2.723	2.698	2.720	2.915	2.861	2.796	3.234	3.183	3.105
2	2.554	2.530	2.517	2.657	2.599	2.599	2.754	2.734	2.746
5	2.333	2.326	2.337	2.344	2.339	2.338	2.206	2.269	2.286
10	2.129	2.157	2.168	2.045	2.103	2.119	1.745	1.862	1.888
20	1.982	1.986	2.031	1.751	1.821	1.901	1.304	1.425	1.478

The table displays the 95 percentile from the empirical distribution of the statistic $t/\sqrt{K}$ when $K = \left[T^{\delta}\right]$. The values of the non-centrality parameter c are displayed in the first column. The forecasting horizon is determined as $K = \left[T^{\delta}\right]$, for the values $\{T, \delta\}$ in the table. The simulations are based on 25,000 replications from the data generating process as described in the text.

Table 2: Empirical Percentile $K = [\delta T]$

$$\Pr\left(t/\sqrt{K} \leq \xi\right) = 0.95$$

Correlation $\gamma = 0$

	$\delta = 0.25$			$\delta = 0.50$			$\delta = 0.75$		
c	T=200	T=500	T=1000	T=200	T=500	T=1000	T=200	T=500	T=1000
1/2	1.662	1.653	1.647	1.170	1.181	1.190	0.673	0.688	0.703
2	1.543	1.581	1.558	1.097	1.100	1.115	0.660	0.664	0.672
5	1.395	1.390	1.413	0.987	0.997	0.998	0.611	0.629	0.648
10	1.218	1.196	1.216	0.827	0.830	0.848	0.563	0.566	0.562
20	0.946	0.955	0.930	0.646	0.661	0.650	0.478	0.490	0.484

Correlation $\gamma = -0.90$

	$\delta = 0.25$			$\delta = 0.50$			$\delta = 0.75$		
c	T=200	T=500	T=1000	T=200	T=500	T=1000	T=200	T=500	T=1000
1/2	3.252	3.237	3.223	2.917	2.984	2.948	1.714	1.705	1.719
2	2.737	2.774	2.796	2.595	2.619	2.642	1.626	1.628	1.623
5	2.214	2.233	2.258	2.096	2.124	2.124	1.424	1.456	1.444
10	1.767	1.803	1.794	1.614	1.644	1.616	1.188	1.225	1.217
20	1.337	1.356	1.361	1.133	1.155	1.167	0.906	0.930	0.929

The table displays the 95 percentile from the empirical distribution of the statistic $t/\sqrt{K}$ when $K = [\delta T]$. The values of the non-centrality parameter c are displayed in the first column. The forecasting horizon is determined as $K = [\delta T]$, for the values $\{T, \delta\}$ in the table. Simulation is based on 25,000 replications from the data generating process as described in the text.

Table 3: Simulated Asymptotic 95 Percentiles of $t/\sqrt{K}$ (5% Critical Values)

	Correlation $\gamma = 0$											
	c=1/2				c=5				c=20			
Percentile	*10%*	*5%*	*2.5%*	*1%*	*10%*	*5%*	*2.5%*	*1%*	*10%*	*5%*	*2.5%*	*1%*
$[\kappa T]$												
K=12	1.286	1.659	1.969	2.337	1.260	1.609	1.951	2.324	1.205	1.546	1.866	2.233
K=24	1.280	1.666	1.993	2.380	1.254	1.636	1.968	2.376	1.131	1.467	1.754	2.095
K=48	1.304	1.695	2.029	2.522	1.226	1.597	1.921	2.291	0.996	1.295	1.557	1.883
K=60	1.295	1.711	2.061	2.458	1.205	1.574	1.925	2.351	0.946	1.237	1.492	1.778
$[T^{\alpha}]$	1.282	1.645	1.960	2.326	1.282	1.645	1.960	2.326	1.282	1.645	1.960	2.326
	Correlation $\gamma = -0.90$											
	c=1/2				c=5				c=20			
Percentile	*10%*	*5%*	*2.5%*	*1%*	*10%*	*5%*	*2.5%*	*1%*	*10%*	*5%*	*2.5%*	*1%*
$[\kappa T]$												
K=12	2.455	2.786	3.072	3.411	1.992	2.325	2.637	3.004	1.592	1.920	2.212	2.564
K=24	2.492	2.855	3.180	3.530	1.995	2.347	2.667	3.069	1.528	1.853	2.156	2.502
K=48	2.600	2.988	3.318	3.753	1.967	2.336	2.659	3.042	1.416	1.700	1.968	2.271
K=60	2.650	3.050	3.400	3.894	1.977	2.348	2.667	3.103	1.362	1.633	1.870	2.148
$[T^{\alpha}]$	2.405	2.702	2.998	3.311	1.999	2.350	2.647	2.991	1.689	2.061	2.353	2.719

The table shows the asymptotic percentiles used to compute the empirical sizes in Tables 4 and 5 for both $K = [\kappa T]$ and $K = [T^{\alpha}]$, where the values of (κ, α) are implied for the values of K in the table.

Table 4: Empirical Sizes of $t/\sqrt{K},\ K = [T^{\alpha}]$

			Correlation $\gamma = 0$						Correlation $\gamma = -0.90$				
K=12	Nom. Size	c	1/2	2	5	10	20	c	1/2	2	5	10	20
	0.100		0.101	0.100	0.096	0.092	0.084		0.107	0.100	0.100	0.090	0.084
	0.050		0.050	0.052	0.048	0.046	0.040		0.058	0.520	0.050	0.042	0.038
	0.025		0.026	0.026	0.023	0.023	0.018		0.030	0.029	0.025	0.021	0.019
	0.010		0.011	0.011	0.010	0.009	0.007		0.014	0.013	0.010	0.008	0.007
K=24													
	0.100		0.100	0.099	0.095	0.086	0.074		0.116	0.102	0.097	0.087	0.069
	0.050		0.051	0.049	0.047	0.040	0.032		0.067	0.056	0.050	0.040	0.028
	0.025		0.027	0.026	0.025	0.019	0.013		0.035	0.033	0.027	0.020	0.014
	0.010		0.013	0.011	0.011	0.007	0.004		0.018	0.016	0.012	0.008	0.005
K=48													
	0.100		0.100	0.095	0.089	0.072	0.049		0.134	0.111	0.097	0.077	0.047
	0.050		0.052	0.051	0.045	0.032	0.020		0.082	0.064	0.050	0.033	0.017
	0.025		0.028	0.027	0.022	0.015	0.007		0.049	0.039	0.027	0.016	0.007
	0.010		0.013	0.012	0.009	0.006	0.002		0.027	0.019	0.013	0.007	0.002
K=60													
	0.100		0.103	0.093	0.088	0.072	0.042		0.140	0.116	0.096	0.069	0.040
	0.050		0.055	0.050	0.044	0.033	0.015		0.087	0.068	0.049	0.028	0.013
	0.025		0.029	0.027	0.023	0.017	0.005		0.052	0.040	0.025	0.013	0.005
	0.010		0.014	0.013	0.010	0.006	0.001		0.030	0.021	0.012	0.005	0.001

The table displays the empirical sizes at the nominal levels in the first column when using asymptotic critical values of $t/\sqrt{K}$ for a sample of 500 observations.The values of the non-centrality parameter c are displayed in the first column. The forecasting horizon is determined as $K = [T^{\alpha}]$ for the values of K in the table. Simulation is based on 25,000 replications from the data generating process as described in the text.

Table 5: Empirical Sizes of $t/\sqrt{K},\ K = [\kappa T]$

		c	Correlation $\gamma = 0$					c	Correlation $\gamma = -0.90$				
K=12	Nom. Size	c	1/2	2	5	10	20	c	1/2	2	5	10	20
	0.100		0.100	0.102	0.103	0.095	0.098		0.097	0.099	0.100	0.103	0.097
	0.050		0.052	0.049	0.053	0.048	0.049		0.048	0.050	0.049	0.050	0.051
	0.025		0.026	0.025	0.025	0.024	0.023		0.024	0.027	0.023	0.026	0.027
	0.010		0.011	0.010	0.011	0.009	0.009		0.010	0.011	0.009	0.011	0.011
K=24													
	0.100		0.097	0.103	0.101	0.098	0.100		0.099	0.098	0.098	0.096	0.095
	0.050		0.050	0.054	0.050	0.050	0.049		0.050	0.049	0.050	0.048	0.046
	0.025		0.025	0.027	0.026	0.025	0.026		0.024	0.024	0.026	0.025	0.021
	0.010		0.010	0.010	0.010	0.011	0.011		0.010	0.011	0.010	0.009	0.009
K=48													
	0.100		0.097	0.099	0.010	0.097	0.101		0.098	0.097	0.102	0.096	0.094
	0.050		0.050	0.049	0.049	0.050	0.050		0.049	0.048	0.050	0.048	0.047
	0.025		0.026	0.025	0.024	0.023	0.025		0.026	0.023	0.025	0.023	0.022
	0.010		0.010	0.010	0.011	0.008	0.008		0.011	0.010	0.010	0.009	0.009
K=60													
	0.100		0.097	0.097	0.101	0.098	0.099		0.093	0.098	0.100	0.099	0.091
	0.050		0.049	0.048	0.049	0.048	0.049		0.046	0.048	0.048	0.049	0.045
	0.025		0.023	0.025	0.023	0.024	0.025		0.024	0.024	0.025	0.025	0.022
	0.010		0.017	0.011	0.008	0.009	0.010		0.010	0.011	0.010	0.010	0.009

The table displays the empirical sizes at the nominal levels in the first column when using asymptotic critical values of $t/\sqrt{K}$ for a sample of 500 observations.The values of the non-centrality parameter c are displayed in the first column. The forecasting horizon is determined as $K = [\kappa T]$ for the values of K in the table. Simulation is based on 25,000 replications from the data generating process as described in the text.

Table 6: Descriptive Statistics of REIT Data

	1972-2003				1972-1989				1990-2003			
	Mean	S.D.	$\hat{\rho}$	ADF	Mean	S.D	$\hat{\rho}$	ADF	Mean	S.D.	$\hat{\rho}$	ADF
						Returns						
AREIT	10.80	15.17	0.08	-13.69	9.95	17.25	0.05	-10.88	11.91	11.98	0.16^b	-7.74
EREIT	13.10	13.46	0.11^b	-13.65	13.53	14.28	0.12^b	-10.82	12.55	12.36	0.10	-8.07
MREIT	9.22	19.93	0.04	-12.78	6.12	20.47	0.00	-11.06	13.25	19.22	0.10	-6.93
HREIT	10.14	19.32	0.01	-13.32	9.88	21.59	-0.07	-10.81	10.50	15.96	0.24^a	-7.59
						Dividend-Price Ratios						
AREIT	1.36	0.19	0.996	-2.68	1.53	0.23	0.995	-2.46	1.14	0.10	0.996	-2.31
EREIT	0.66	0.13	0.998	-2.00	0.87	0.14	0.998	-2.05	0.39	0.03	0.993	-2.44
MREIT	4.91	0.85	0.997	-1.96	2.89	0.35	0.994	-2.73	7.53	0.70	0.993	-1.65
HREIT	1.84	0.33	0.997	-2.12	1.44	0.24	0.996	-2.05	2.35	0.38	0.996	-1.67
Obs.	384				217				167			

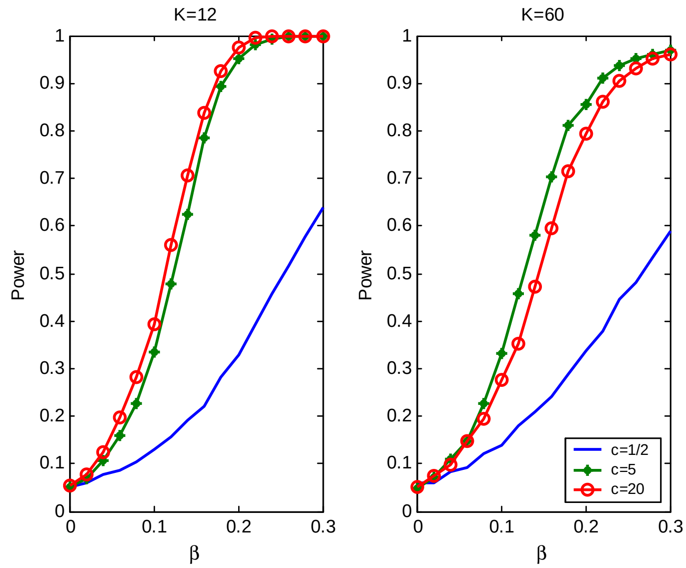
All variables are available at monthly frequency and are displayed in annualized percentage points (mean and standard deviation). Returns and dividend-price ratios are presented for four portfolios: (1) AREIT–all REITs in the NAREIT index; (2) EREIT–Equity REITs only; (3) MREIT–Mortgage REITs only; (4) HREIT–Hybrid REITs only. The dividend-price ratios are smoothed with a 12-month moving average, as in Fama and French (1988), to decrease any seasonal effects in the series. The column $\hat{\rho}$ denotes the autoregressive coefficient of the series, and ADF is the Augmented Dickey-Fuller test for the null of unit root, where the auxiliary regression includes a constant. The ADF test is specified with 1 lag for returns and with 12 lags for dividend-price ratios. The asymptotic critical values at the [1%, 5%, 10%] level are respectively [-3.45, -2.87,-2.57]. The dividend-price ratios are very serially correlated and the ADF test cannot reject the null of unit root for most of the portfolios in the samples. For returns, significance in the autocorrelation coefficient is denoted as a (1%), b (5%) and c (10%).

Table 7: Long-Horizon Predictive Regressions

K	Period 1972-2003 $\hat{\beta}_K$	t	t^{NW}	$t/\sqrt{K}$	$\mathcal{R}^2_K$	Period 1972-1989 $\hat{\beta}_K$	t	t^{NW}	$t/\sqrt{K}$	$\mathcal{R}^2_K$	Period 1990-2003 $\hat{\beta}_K$	t	t^{NW}	$t/\sqrt{K}$	$\mathcal{R}^2_K$
AREIT															
1	0.01	1.72	1.35	$1.72^{C,\mathcal{C}}$	0.01	0.01	1.28	1.15	1.28	0.01	0.02	1.76	1.56	1.76^{C}	0.02
6	0.06	4.80	2.61	$1.96^{C,\mathcal{C}}$	0.06	0.06	3.58	2.17	1.46^{C}	0.06	0.12	4.65	2.25	1.90^{C}	0.12
12	0.15	8.06	3.48	$2.33^{B,\mathcal{B}}$	0.15	0.14	6.04	2.65	$1.75^{C,\mathcal{C}}$	0.15	0.26	8.72	3.13	$2.52^{B,\mathcal{B}}$	0.33
36	0.45	18.42	8.29	$3.07^{A,\mathcal{A}}$	0.50	0.46	15.93	7.59	$2.65^{A,\mathcal{B}}$	0.59	0.35	7.35	2.88	1.23	0.30
60	0.50	19.95	5.97	$2.58^{A,\mathcal{B}}$	0.55	0.44	14.54	12.76	$1.88^{C,\mathcal{C}}$	0.58	0.40	10.85	7.60	1.40^{C}	0.53
EREIT															
1	0.00	1.18	1.07	1.18	0.00	0.00	0.74	0.72	0.75	0.00	0.02	1.79	1.59	1.71^{C}	0.02
6	0.03	3.20	1.50	1.31	0.03	0.02	1.73	1.00	0.70	0.01	0.12	4.82	2.36	$1.97^{C,\mathcal{C}}$	0.13
12	0.08	6.11	2.41	$1.76^{C,\mathcal{C}}$	0.09	0.08	4.03	2.46	1.16	0.07	0.27	9.48	3.50	$2.74^{A,\mathcal{B}}$	0.37
36	0.26	14.18	6.33	$2.36^{B,\mathcal{B}}$	0.37	0.30	10.37	6.09	$1.73^{C,\mathcal{C}}$	0.38	0.36	8.08	3.87	1.35	0.34
60	0.39	22.43	7.87	$2.90^{A,\mathcal{A}}$	0.61	0.33	11.29	16.30	$1.46^{C,\mathcal{C}}$	0.45	0.42	11.61	8.44	1.50^{C}	0.56
MREIT															
1	0.01	2.87	3.43	$2.87^{A,\mathcal{A}}$	0.02	0.01	1.54	1.64	1.54^{C}	0.01	0.00	3.92	3.50	$3.92^{A,\mathcal{A}}$	0.09
6	0.07	6.92	2.89	$2.83^{A,\mathcal{A}}$	0.11	0.07	4.51	1.79	$1.84^{C,\mathcal{C}}$	0.09	0.33	10.26	6.09	$4.19^{A,\mathcal{A}}$	0.40
12	0.16	9.61	2.76	$2.77^{A,\mathcal{A}}$	0.20	0.16	7.33	1.71	$2.12^{C,\mathcal{C}}$	0.21	0.67	16.14	8.15	$4.66^{A,\mathcal{A}}$	0.63
36	0.35	12.48	6.61	$2.08^{C,\mathcal{C}}$	0.31	0.46	18.93	7.97	$3.15^{A,\mathcal{A}}$	0.67	1.38	15.13	12.54	$2.52^{C,\mathcal{C}}$	0.64
60	0.18	6.15	3.68	0.79	0.10	0.36	11.57	10.61	1.49^{C}	0.46	0.09	0.77	0.84	0.10	0.01
HREIT															
1	0.01	1.75	1.76	$1.75^{C,\mathcal{C}}$	0.01	0.01	0.98	0.84	0.98	0.00	0.01	1.84	1.72	1.84^{C}	0.02
6	0.06	4.96	2.98	$2.02^{C,\mathcal{C}}$	0.06	0.06	3.05	2.31	1.25	0.04	0.10	4.47	2.49	1.83^{C}	0.11
12	0.15	7.75	3.56	$2.24^{B,\mathcal{C}}$	0.14	0.13	5.08	3.08	1.47^{C}	0.11	0.22	6.70	2.97	1.93^{C}	0.23
36	0.51	15.35	7.69	$2.56^{A,\mathcal{B}}$	0.40	0.48	13.35	5.57	$2.23^{B,\mathcal{C}}$	0.50	0.72	11.87	3.79	$1.98^{C,\mathcal{C}}$	0.52
60	0.67	15.76	4.58	$2.03^{B,\mathcal{C}}$	0.43	0.53	13.90	10.08	$1.79^{B,\mathcal{C}}$	0.55	1.18	11.51	13.06	1.49^{C}	0.56

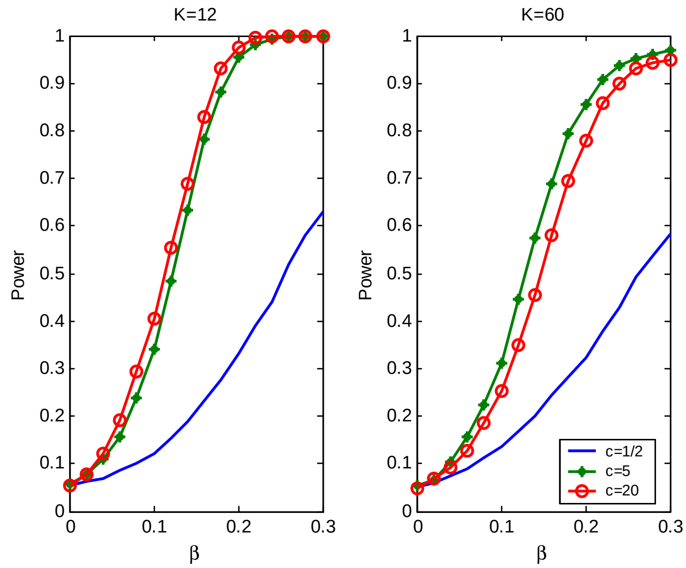
Results from forecasting REIT returns using a lagged dividend-price ratio for: (1) AREIT–all REITs in the NAREIT index; (2) EREIT–Equity REITs; (3) MREIT–Mortgage REITs; (4) HREIT–Hybrid REITs. Returns and dividend-price ratios are expressed in logs. Long-horizon returns are determined as a rolling sum from period t to $t+K$ of one-period log returns. The first column displays the OLS estimate of the regression slope ($\hat{\beta}_K$). The second, third and fourth columns display the OLS t-statistic (t), the Newey-West (1987) adjusted t-statistic using truncation lag $\max\{1, K-1\}$, (t^{NW}), and the $\sqrt{K}$-normalized t-statistic $\left(t/\sqrt{K}\right)$. The fifth column shows the $\mathcal{R}^2$ of the regression. The $t/\sqrt{K}$ test is evaluated at the conventional 1%, 5% and 10% statistical levels by using the estimator of c in [Valkanov (2003)] and {Phillips *et al.* (2001)}. Significance is denoted, respectively, as $[A]\{\mathcal{A}\}$, $[B]\{\mathcal{B}\}$, and $[C]\{\mathcal{C}\}$.

Figure 1: Power of $t/\sqrt{K}$ for Various Values of c, $K = [T^{\alpha}]$



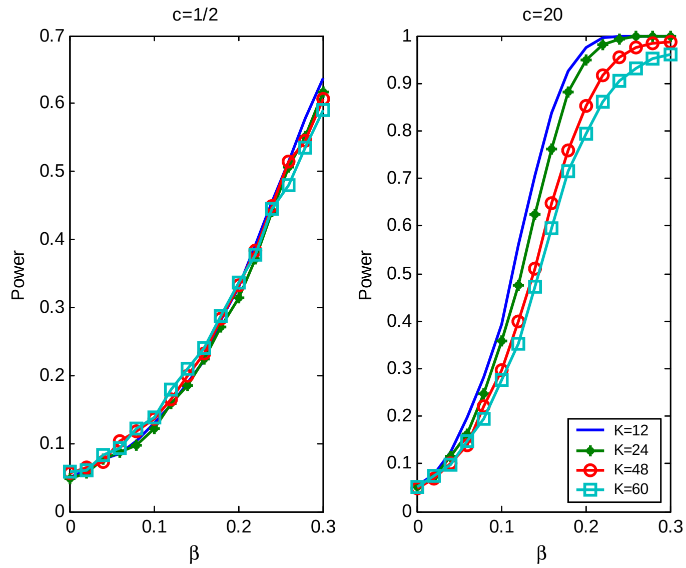
The figure shows the power of the $t/\sqrt{K}$-statistic as a function of the parameter $c = \{1/2,\ 5,\ 20\}$ for a nominal size of 5%. The sample size is 500 and the data generating process is simulated for various values of β, and for $\gamma = -0.90$, $K = 12$, and $K = 60$. The asymptotic critical values are obtained from the theory $K = [T^{\alpha}]$.

Figure 2: Power of $t/\sqrt{K}$ for Various Values of c, $K = [\kappa T]$



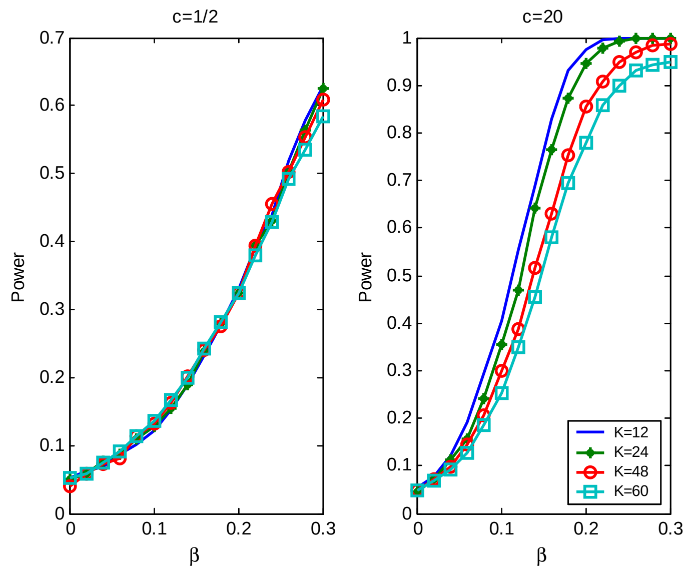
The figure shows the power of the $t/\sqrt{K}$-statistic as a function of the parameter $c = \{1/2,\ 5,\ 20\}$ for a nominal size of 5%. The sample size is 500 and the data generating process is simulated for various values of β, and for $\gamma = -0.90$, $K = 12$, and $K = 60$ as described in the text. The asymptotic critical values are obtained from the theory $K = [\kappa T]$.

Figure 3: Power of $t/\sqrt{K}$ at Various Horizons, $K = [T^{\alpha}]$



The figure shows the power of the $t/\sqrt{K}$-statistic as a function of the horizon K for a nominal size of 5%. The sample size is 500 and the data generating process is simulated for various values of β, and for $\gamma = -0.90$, $c = \{1/2, 20\}$, $K = \{12,\ 24,\ 48,\ 60\}$ as described in the text. The asymptotic critical values are obtained from the theory $K = [T^{\alpha}]$.

Figure 4: Power of $t/\sqrt{K}$ at Various Horizons, $K = [\kappa T]$



The figure shows the power of the $t/\sqrt{K}$-statistic as a function of the horizon K for a nominal size of 5%. The sample size is 500 and the data generating process is simulated for various values of β, and for $\gamma = -0.90$, $c = \{1/2, 20\}$, $K = \{12,\ 24,\ 48,\ 60\}$ as described in the text. The asymptotic critical values are obtained from the theory $K = [\kappa T]$.