ELSEVIER

Journal of Empirical Finance 8 (2001) 459–491

Journal of EMPIRICAL FINANCE

www.elsevier.com/locate/econbase

Why long horizons? A study of power against persistent alternatives ☆

John Y. Campbell *

Department of Economics, Harvard University, Littauer Center, 1875 Cambridge Street, Cambridge, MA, 02138, USA

Abstract

This paper studies tests of predictability in regressions with a given AR(1) regressor and an asset return dependent variable measured over a short or long horizon. The paper shows that when there is a persistent predictable component in the return, an increase in the horizon may increase the R^2 statistic of the regression and the approximate slope of a predictability test. Monte Carlo experiments show that long-horizon regression tests have serious size distortions when asymptotic critical values are used, but some versions of such tests have power advantages remaining after size is corrected.

JEL classification: C22; G12

Keywords: Long-horizon regressions; Power; Approximate slope

1. Introduction

During the last 20 years financial econometricians have become accustomed to running long-horizon regressions. In these regressions the dependent variable is an

☆ Earlier versions of this paper carried the title "Why Long Horizons? Asymptotic Power Against Persistent Alternatives". I am grateful to the National Science Foundation and the Houblon–Norman Fellowship at the Bank of England for financial support, and to Martin Lettau for able research assistance. John Cochrane, Frank Diebold, Rob Engle, Robert Hodrick, and James Stock made helpful comments on earlier drafts.

* Fax: +1-617-495-8570.

E-mail address: John_Campbell@harvard.edu (J.Y. Campbell).

0927-5398/01/$ - see front matter
PII: S0927-5398(01)00037-8

asset return measured over a longer time period than the sampling interval. Thus with data sampled monthly, the dependent variable might be a return measured over one or even several years. The discrepancy between the return horizon and the sampling interval leads to serial correlation in the regression error even under the null hypothesis that the asset return is uncorrelated with the regressors. This is handled using the now familiar asymptotic theory of Hansen (1982) and White (1984).

Long-horizon regressions were originally introduced for situations where asset returns are not observed over short horizons. Thus, Hansen and Hodrick (1980) studied forward exchange rates and could only measure returns over a three-month horizon. Long-horizon regressions allowed them to use all their weekly data rather than sampling the data quarterly.

More recently, long-horizon regressions have been used even when short-horizon returns are observable. Fama and French (1988a) regressed stock returns on lagged stock returns, measuring returns at horizons from 1 to 10 years, and found stronger evidence of predictability at longer horizons. This regression, and the related variance ratio tests of Cochrane (1988), Lo and MacKinlay (1988), and Poterba and Summers (1988), have the feature that the choice of horizon is bound up with the choice of forecasting variable. Other applications use the same forecasting variable at all horizons. Fama and French (1988b, 1989) and Campbell and Shiller (1988b), for example, regress stock returns on dividend-price and earnings-price ratios and interest rates, again finding the strongest evidence of predictability at longer horizons. Volatility tests, which have generated striking results since their introduction by LeRoy and Porter (1981) and Shiller (1981), can be interpreted as regressions of weighted long-horizon stock returns onto stock prices or stationary transformations of stock prices (Campbell and Shiller, 1988b; Campbell et al., 1997; Scott, 1985; Shiller, 1989). Macroeconomists have used long-horizon regressions to detect predictable components in nominal interest rates, real interest rates, and inflation (Campbell and Shiller, 1991; Fama, 1990; Fama and Bliss, 1987; Mishkin, 1990a,b), and more recently in exchange rates (Mark, 1995).

These findings suggest that in finite samples, tests of predictability in long-horizon regressions must have either greater power than short-horizon regression tests, or more serious size distortions, or perhaps both. Power and size have been explored for the special case in which returns are regressed on lagged returns (Faust, 1992; Lo and Mackinlay, 1989; Richardson and Smith, 1991a,b; Richardson and Stock, 1989), but this case makes it hard to distinguish clearly the effects of changing the horizon from the effects of changing the variable used to forecast returns.

In this paper I study the case in which returns are regressed onto a forecasting variable, which remains the same no matter what horizon is used. Specifically, I construct an example in which the forecasting variable captures all predictability in one-period returns and follows an AR(1) process so that it is the optimal

forecaster of returns at all horizons.[1] I use this example to explore the properties of standard long-horizon regressions, and also of weighted long-horizon regressions, in which more distant future returns receive geometrically declining rather than equal weights in forming the long-horizon return. Volatility tests are closely related to long-horizon regressions weighted in this way.

The first result is that the regression R^2 statistic can increase with the return horizon when the forecasting variable is variable and persistent. Of course, this does not mean that long-horizon regression tests have statistical advantages. Indeed, at first sight it seems implausible that there could be any statistical reason to lengthen the horizon in the example studied here. The return and the forecasting variable form a restricted first-order vector autoregression, and maximum likelihood estimates of the coefficients are obtainable by single-period OLS regressions. The standard analysis of power against a sequence of local alternatives, as summarized in Engle (1984), then implies that a one-period regression Wald test is asymptotically optimal.

A different asymptotic power analysis, however, gives a different result. Under a fixed alternative model within the class studied here it is straightforward to calculate approximate slope, a measure of asymptotic power due to Bahadur (1960) and Geweke (1981), for regressions with any horizon. Approximate slope can be much higher for long-horizon regressions than for short-horizon regressions when the AR(1) process for the expected return is persistent, and when the correlation between unexpected return and the innovation in the expected return is negative. I use the loglinear approximate asset pricing framework of Campbell and Shiller (1988a,b) and Campbell (1991) to show that such a negative correlation is plausible, especially when the expected asset return is persistent.

The approximate slope advantages of long-horizon regression require that innovations in asset returns are negatively correlated with future returns, but they do not require negative serial correlation in realized asset returns. Returns can be white noise and yet have a persistent predictable component which increases the approximate slope of long-horizon regression tests. In the U.S. stock market, for example, long-horizon regressions can have approximate slope advantages whether or not there is mean-reversion in returns, as debated by Fama and French (1988a), Jegadeesh (1991), Poterba and Summers (1988), and Richardson and Smith (1991b) among others.

I also consider some variations of the basic weighted or unweighted long-horizon regression. One variation, sometimes used by Fama and French (1988b, 1989), is to sample the data only once every K periods, where K is the horizon of the regression. This simplifies inference by eliminating the serial correlation of the

[1] Mankiw and Shapiro (1986) and Stambaugh (1999) have used this example to study the finite-sample size of one-period regressions, while Nelson and Kim (1993) have used it to study the finite-sample size of long-horizon regressions. The question of power has received much less attention. Hodrick (1992) uses a related but more complicated VAR model to study both size and power.

regression error, but its approximate slope in the AR(1) example is usually smaller, and never much greater, than the approximate slope of a short-horizon regression.

A second variation, discussed by Cochrane (1991), Hodrick (1992), and Jegadeesh (1991), is to run a short-horizon regression in which the regressor is backward-averaged over a long period. The numerator of the coefficient in this regression is the same as the numerator of a long-horizon regression coefficient. Nevertheless, the backward-averaged regression has smaller approximate slope than a basic short-horizon regression in the AR(1) example of this paper.

A third variation, suggested by Hodrick (1992) and Bollerslev and Hodrick (1992), is to run a long-horizon regression but to calculate the asymptotic standard error of the regression coefficient imposing the null hypothesis that returns are unpredictable.[2] This also eliminates the approximate slope advantages of the basic long-horizon regression. If the alternative hypothesis is true, then there is less uncertainty at long horizons than one would think from looking only at short horizons; approximate slope increases with the horizon only when one allows the data to reveal this fact.

Of course, approximate slope is an asymptotic concept which is only useful if it accurately predicts performance in finite samples. The paper runs some Monte Carlo experiments to study this question. The Monte Carlo work here differs in two ways from previous Monte Carlo studies such as those of Bollerslev and Hodrick (1992), Campbell (1991), Goetzmann and Jorion (1993), Hodrick (1992), and Nelson and Kim (1993). First, most studies in the literature use a single data generating process estimated from or calibrated to fit U.S. stock return data. The simple example of this paper allows me to vary the parameters of the model and to compare the asymptotic and finite-sample properties of long-horizon regressions. Second, most studies concentrate on size whereas the emphasis here is on the power of long-horizon regressions.[3]

The organization of the paper is as follows. Section 2 develops the example that is used to structure the investigation. Section 3 compares the R^2 statistics and approximate slopes of short- and long-horizon regressions. Section 4 gives Monte Carlo results, and Section 5 concludes.

2. A simple model with time-varying expected returns

The basic model considered here ignores constant terms for simplicity and writes the log stock return r_{t+1} as a coefficient λ times an observable zero-mean

[2] Jegadeesh (1991) and Richardson and Smith (1991b) also impose the null hypothesis when calculating standard errors for long-horizon regressions of returns on lagged returns. Bollerslev and Hodrick (1992) credit Lars Hansen with the basic idea.

[3] Bollerslev and Hodrick (1992) and Hodrick (1992) also have some discussion of power.

variable x_t plus a random error v_{t+1}:

$$r_{t+1} = \lambda x_t + v_{t+1}. \tag{2.1}$$

The variable x_t is assumed to follow an AR(1) process,

$$x_{t+1} = \phi x_t + u_{t+1}, \qquad -1 < \phi < 1. \tag{2.2}$$

When the AR coefficient ϕ is close to one, the x_t process is highly *persistent*. The shocks v_{t+1} and u_{t+1} form a serially uncorrelated vector but they can be contemporaneously correlated, and the case of negative correlation is of special interest here.

Eqs. (2.1) and (2.2) imply that the univariate process for the log stock return is an ARMA (1,1) with autoregressive coefficient ϕ. The bivariate process for the stock return and the forecasting variable is a first-order vector autoregression with zero coefficients on the lagged stock return. This is a standard model in which Ordinary Least Squares estimates of Eqs. (2.1) and (2.2) are equivalent to Seemingly Unrelated Regressions estimates (since the same explanatory variable appears in both equations).

The model Eqs. (2.1) and (2.2) has five parameters: the coefficients λ and ϕ, and the variances and covariances σ_u^2, σ_v^2, and σ_{uv}. Since the units of measurement for r_{t+1} and x_t are arbitrary, up to two of these parameters can be normalized. It is convenient to normalize $\lambda = 1$, so that the variable x_t can be interpreted as the expected stock return. A problem with this normalization is that it fails to represent the case of unpredictable stock returns, $\lambda = 0$; however, this case is approached in the limit when σ_v^2 becomes large relative to σ_u^2.

2.1. A reparameterization

The economics of this example can be understood more clearly, and the example can be parameterized in a more useful way, with the help of the loglinear approximate relation between dividends, prices, and returns proposed by Campbell and Shiller (1988a,b) and Campbell (1991). Using this approximation, it is straightforward to show that the stock price implied by the AR(1) example Eqs. (2.1) and (2.2) is the sum of two terms. The first term is the expected discounted value of future dividends, which is close to a random walk, while the second term is a stationary AR(1) process. This two-component description of stock prices is often found in the literature (Fama and French, 1988a; Poterba and Summers, 1988; Summers, 1986).

More relevant here is the loglinear decomposition for returns derived by Campbell (1991):

$$\begin{aligned} v_{t+1} &\equiv r_{t+1} - E_t r_{t+1} \\ &= (E_{t+1} - E_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}. \end{aligned} \tag{2.3}$$

In this expression the discount factor $\rho \approx 1/(1+D/P)$, where D/P is the mean dividend yield. Eq. (2.3) says that unexpected stock returns must be associated with changes in expectations of future dividends or real returns. An increase in expected future dividends is associated with a capital gain today, while an increase in expected future returns is associated with a capital loss today. The reason is that with a given dividend stream, higher future returns can only be generated by future price appreciation from a lower current price. It is convenient to simplify the notation of Eq. (2.3) to

$$v_{t+1} = v_{\mathrm{d},t+1} - v_{\mathrm{r},t+1}, \tag{2.4}$$

where $v_{\mathrm{d},t+1}$ is the change in expectations of future dividends in Eq. (2.3), and $v_{\mathrm{r},t+1}$ is the change in expectations of future returns.

In the AR(1) example, with λ normalized to one, the general stock return Eq. (2.4) simplifies because the innovation in expected future stock returns, $v_{\mathrm{r},t+1}$, is given by $\rho u_{t+1}/(1-\rho\phi)$. Thus Eq. (2.1) becomes

$$r_{t+1} = x_t + v_{\mathrm{d},t+1} - \frac{\rho u_{t+1}}{1-\rho\phi}. \tag{2.5}$$

Eqs. (2.2) and (2.5) form the reparameterized system. Since u_{t+1} and $v_{\mathrm{d},t+1}$ can be correlated, this reparameterization places no restrictions on the original system Eqs. (2.1) and (2.2).

Eq. (2.5) can be used to calculate the variance of the one-period stock return. Write $\mathrm{Var}(x_{t+1}) = \sigma_x^2$ and $\mathrm{Var}(u_{t+1}) = \sigma_u^2$, and note that from Eq. (2.2) these are related by $\sigma_u^2 = (1-\phi^2)\sigma_x^2$. Write $\mathrm{Var}(v_{\mathrm{d},t+1}) = \sigma_{\mathrm{d}}^2$, so that σ_{d}^2 is the variance of news about all future dividends rather than the variance of the dividend itself. Also write $\mathrm{Cov}(v_{\mathrm{d},t+1}, u_{t+1}) = \mathrm{Cov}(v_{\mathrm{d},t+1}, x_{t+1}) = \sigma_{\mathrm{d}x}$. Then the one-period stock return variance is

$$\mathrm{Var}(r_{t+1}) = \frac{1-2\rho\phi+\rho^2}{(1-\rho\phi)^2}\sigma_x^2 - \frac{2\rho}{(1-\rho\phi)}\sigma_{\mathrm{d}x} + \sigma_{\mathrm{d}}^2. \tag{2.6}$$

This equation simplifies considerably if one uses the approximation $\rho \approx 1$ (which will be accurate provided that $\phi < \rho$). Eq. (2.6) then becomes

$$\mathrm{Var}(r_{t+1}) = \frac{2}{1-\phi}\left(\sigma_x^2 - \sigma_{\mathrm{d}x}\right) + \sigma_{\mathrm{d}}^2 = \frac{2}{1-\phi}\left(1-\beta_{\mathrm{d}x}\right)\sigma_x^2 + \sigma_{\mathrm{d}}^2, \tag{2.7}$$

where $\beta_{\mathrm{d}x} \equiv \sigma_{\mathrm{d}x}/\sigma_x^2$ is the regression coefficient of dividend news on the expected return at time $t+1$. To interpret Eq. (2.7), note that σ_{d}^2, the variance of dividend news, appears with a unit coefficient. σ_x^2, the variance of the expected return, has a coefficient of $2/(1-\phi)$; persistence in the expected return process increases the variability of realized returns, for small but persistent changes in expected returns have large effects on prices and thus on realized returns. Finally $\sigma_{\mathrm{d}x}$, the covariance between dividend news and the expected return, has a

coefficient of $-2/(1-\phi)$; a positive covariance between dividend news and expected returns reduces the variance of the stock return because good dividend news raises the stock price directly but lowers it indirectly through the effect on expected returns. When $\beta_{dx}=1$, the σ_{dx} and σ_x^2 terms exactly offset each other so that the variance of stock returns is just the variance of dividend news σ_d^2.

2.2. Serial correlation and predictability of stock returns

Within the AR(1) example it is straightforward to calculate the autocovariances of realized stock returns:

$$\begin{aligned}\mathrm{Cov}(r_{t+1},r_{t+1+i}) &= \mathrm{Cov}(x_t,r_{t+i}) + \mathrm{Cov}(v_{t+1},r_{t+i})\\ &= \mathrm{Cov}(x_t,x_{t+i}) + \mathrm{Cov}(v_{t+1},x_{t+i})\\ &= \phi^{i-1}\sigma_x^2\left[\phi - \frac{\rho(1-\phi^2)}{1-\rho\phi} + \beta_{dx}\right]. \end{aligned} \tag{2.8}$$

The autocovariances of stock returns all have the same sign and die out geometrically at rate ϕ, consistent with the univariate ARMA(1,1) representation for stock returns. The sign of the autocovariances depends on offsetting effects: the positive autocovariances of expected returns appear in realized returns as well, but a positive innovation to future expected returns causes a contemporaneous capital loss, and this introduces negative autocovariance into realized returns. The latter effect dominates provided that $\phi<\rho$. In addition, a positive covariance between dividend news and expected returns creates positive serial correlation in stock returns because capital gains from good dividend news tend to be followed by predictably higher returns. Overall there is some presumption that changing expected returns create negative autocovariance in realized return unless the covariance between dividend news and expected returns is large and positive.

To see this more clearly, one can simplify Eq. (2.8) by again using the approximation $\rho\approx 1$ to obtain

$$\mathrm{Cov}(r_{t+1},r_{t+1+i}) = -\phi^{i-1}\sigma_x^2(1-\beta_{dx}). \tag{2.9}$$

Eq. (2.9) shows that stock returns will be negatively correlated unless $\beta_{dx}\geq 1$. White noise stock returns are obtained when $\beta_{dx}=1$; in this case stock returns cannot be predicted from their own past history.

Even when stock returns are white noise, they can be predicted if one has knowledge of the expected stock return x_t. Consider the R^2 statistic obtained in a regression of the one-period stock return r_{t+1} on the variable x_t. In population the fitted value is just x_t itself, with variance σ_x^2, while the variance of the return is given by Eq. (2.6) or (2.7) above. Working with Eq. (2.7), the simpler expression that assumes $\rho=1$, the one-period regression R^2 statistic is

$$R^2(1) \equiv \frac{\mathrm{Var}(x_t)}{\mathrm{Var}(r_{t+1})} = \left[\frac{2}{1-\phi}(1-\beta_{dx}) + \frac{\sigma_d^2}{\sigma_x^2}\right]^{-1}. \tag{2.10}$$

As the ratio σ_d^2/σ_x^2 increases, the one-period R^2 statistic declines to zero and stock returns become unpredictable. For any given value of β_{dx} there is a lower bound on the ratio σ_d^2/σ_x^2 because

$$\beta_{dx} \equiv \frac{\sigma_{dx}}{\sigma_x^2} = \frac{\sigma_{du}(1-\phi^2)}{\sigma_u^2} \leq \frac{\sigma_d(1-\phi^2)}{\sigma_u} = \frac{\sigma_d\sqrt{(1-\phi^2)}}{\sigma_x}. \tag{2.11}$$

Hence, $\sigma_d^2/\sigma_x^2 \beta_{dx}^2/(1-\phi^2)$. As σ_d^2/σ_x^2 decreases to this lower bound, the one-period R^2 statistic approaches an upper bound, $R^2_{max}(1)$. When $\beta_{dx}=1$, the lower bound on σ_d^2/σ_x^2 is $1/(1-\phi^2)$ and the upper bound on the R^2 statistic is $1-\phi^2$ under the simplifying assumption that $\rho=1$. When $\beta_{dx}=0$, the lower bound on σ_d^2/σ_x^2 is zero and the upper bound on the R^2 statistic is $(1-\phi)/2$ under the same assumption that $\rho=1$. Thus even in a limiting case in which there is no dividend uncertainty at all, so that the stock is a consol with fixed payoffs and all stock price variation is due to changing expected returns, the one-period R^2 statistic is small when ϕ is large. The reason is that innovations to expected returns cause large unforecastable changes in stock prices when expected returns are persistent. This point is emphasized by Campbell (1991).

Table 1 illustrates the mapping between the properties of stock returns and the parameters of the original model Eqs. (2.1) and (2.2). Panel A of Table 1 sets $\beta_{dx}=0$, while panel B sets β_{dx} at whatever value eliminates the univariate autocorrelations of stock returns. As already discussed, this value is one when $\rho=1$, but the table sets $\rho=0.995$, appropriate for monthly data when the annual dividend yield is roughly 6%. The zero-autocorrelation value of β_{dx} is reported in parentheses in the left column of panel B; it is approximately one for low ϕ, but falls as ϕ approaches one.

Within each panel of Table 1, the rows represent different values of ϕ and the columns represent different values of the one-period R^2 statistic as a fraction of the upper bound $R^2_{max}(1)$.[4] The upper bound itself is reported in the right column of the table. As already discussed, this bound is $(1-\phi)/2$ when $\beta_{dx}=0$ and $\rho=1$, but it is slightly different in panel A of Table 1 because the table sets $\rho=0.995$. However, the exact upper bound does fall with ϕ in the way one would expect from the analysis of the $\rho=1$ case.

For each value of β_{dx}, ϕ, and $R^2(1)/R^2_{max}(1)$, the table reports the implied correlation ρ_{uv} between unexpected stock returns v_{t+1} and innovations in expected stock returns u_{t+1}. In parentheses, the table also reports the relative standard deviations of these two innovations, σ_u/σ_v. The main interest is in the

[4] When $\beta_{dx}=0$ and $\rho=1$, the fraction of $R^2_{max}(1)$ equals one minus the limiting variance ratio for stock returns. Thus in panel A the columns of Table 1 also represent the long-run univariate mean reversion of stock returns. In panel B, of course, there is no univariate mean reversion at all because stock returns are univariate white noise by construction.

Table 1
Implied moments of VAR innovations ρ_{uv}, (σ_u / σ_v)

A: $\beta_{dx} = 0$

$R^2(1)/R^2_{max}(1)$ ϕ	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(1)$
0.100	−0.239 (0.217)	−0.393 (0.355)	−0.595 (0.538)	−0.780 (0.706)	−0.912 (0.825)	−0.955 (0.864)	−1.00 (0.905)	0.453
0.250	−0.254 (0.192)	−0.414 (0.313)	−0.619 (0.468)	−0.799 (0.604)	−0.921 (0.695)	−0.960 (0.725)	−1.00 (0.755)	0.378
0.500	−0.277 (0.140)	−0.446 (0.225)	−0.654 (0.330)	−0.825 (0.416)	−0.933 (0.471)	−0.966 (0.488)	−1.00 (0.505)	0.254
0.750	−0.297 (0.076)	−0.474 (0.121)	−0.682 (0.174)	−0.844 (0.215)	−0.942 (0.240)	−0.971 (0.248)	−1.00 (0.255)	0.129
0.900	−0.308 (0.032)	−0.489 (0.051)	−0.697 (0.073)	−0.854 (0.090)	−0.946 (0.099)	−0.973 (0.102)	−1.00 (0.105)	0.055
0.950	−0.312 (0.017)	−0.494 (0.027)	−0.702 (0.039)	−0.857 (0.047)	−0.947 (0.052)	−0.974 (0.054)	−1.00 (0.055)	0.030
0.980	−0.314 (0.008)	−0.497 (0.012)	−0.704 (0.018)	−0.858 (0.022)	−0.948 (0.024)	−0.974 (0.024)	−1.00 (0.025)	0.016

B: Zero-autocorrelation β_{dx}

$R^2(1)/R^2_{max}(1)$ ϕ (β_{dx})	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(1)$
0.100 (0.994)	−0.032 (0.312)	−0.049 (0.489)	−0.069 (0.684)	−0.083 (0.823)	−0.091 (0.901)	−0.093 (0.923)	−0.096 (0.945)	0.990
0.250 (0.992)	−0.078 (0.292)	−0.121 (0.452)	−0.165 (0.619)	−0.195 (0.731)	−0.211 (0.791)	−0.215 (0.808)	−0.220 (0.824)	0.937
0.500 (0.985)	−0.153 (0.229)	−0.230 (0.345)	−0.304 (0.455)	−0.348 (0.522)	−0.370 (0.555)	−0.376 (0.564)	−0.382 (0.573)	0.750
0.750 (0.966)	−0.224 (0.131)	−0.328 (0.192)	−0.418 (0.244)	−0.468 (0.273)	−0.491 (0.287)	−0.498 (0.290)	−0.503 (0.294)	0.437
0.900 (0.909)	−0.266 (0.056)	−0.385 (0.081)	−0.484 (0.102)	−0.537 (0.113)	−0.562 (0.119)	−0.568 (0.120)	−0.574 (0.121)	0.190
0.950 (0.822)	−0.282 (0.029)	−0.410 (0.042)	−0.518 (0.053)	−0.577 (0.059)	−0.604 (0.062)	−0.612 (0.063)	−0.619 (0.064)	0.098
0.980 (0.602)	−0.297 (0.012)	−0.442 (0.018)	−0.574 (0.023)	−0.651 (0.026)	−0.689 (0.028)	−0.699 (0.028)	−0.708 (0.029)	0.040

lower right part of the table, where the expected stock return is persistent (so ϕ is large) and the predictability of stock returns is close to its upper bound (but small in absolute terms because the upper bound is small when ϕ is large). In this part of the table v_{t+1} and u_{t+1} are highly negatively correlated, and u_{t+1} has a small standard deviation relative to v_{t+1}.

3. Some asymptotic properties of long-horizon regressions

A standard long-horizon regression with horizon K takes the form

$$r_{t+1} + \cdots + r_{t+K} = \beta(K)\, x_t + e_{t+K,K}. \tag{3.1}$$

In the AR(1) example, with the coefficient λ normalized to one, $\beta(K) = (1 + \phi + \ldots + \phi^{K-1}) = (1 - \phi^K)/(1 - \phi)$. The error term $e_{t+K,K}$ depends on shocks arriving between periods $t+1$ and $t+(K-1)$; thus it may be correlated with errors dated $t-(K-1)$ through $t+(K-1)$. The error term is serially uncorrelated if the regression includes data sampled only every K periods, but is serially correlated when all the data are used. I give most attention to the *overlapping* regression that uses all the data, but below I also consider the *non-overlapping* regression that includes only one in K observations.

The standard long-horizon regression weights each one-period return equally in forming a K-period return. An interesting alternative is a regression in which distant future returns receive less weight than near future returns in forming the dependent variable. For simplicity, I consider a regression with geometrically declining weights and an infinite horizon:

$$r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots = \beta^*(\gamma)\, x_t + e_t^*(\gamma), \qquad 0 \le \gamma < 1. \tag{3.2}$$

In a finite sample, of course, the horizon cannot be longer than the remainder of the sample period. But an infinite horizon simplifies the algebra considerably and is practically equivalent to a finite horizon unless the discount factor γ is very close to one. Note that when $\gamma = 0$ (3.2) is just a standard one-period regression. In the AR(1) example the coefficient $\beta^*(\gamma) = 1/(1 - \gamma\phi)$.

The geometrically declining weights in Eq. (3.2) are also of interest because they relate to the large literature on volatility tests. Campbell and Shiller (1988b) and Campbell et al. (1997) show that a volatility test can be interpreted as a restriction on the regression coefficient $\beta^*(\gamma)$ in Eq. (3.2), where the discount factor γ equals ρ and therefore downweights future returns only very gradually.

The remainder of this section explores the way in which the asymptotic properties of long-horizon regressions depend on the horizon K in Eq. (3.1) or the discount factor γ in Eq. (3.2).

3.1. R^2 statistics

Asymptotically, the R^2 statistic for the standard K-period regression (with either overlapping or non-overlapping data) is given by

$$R^2(K) \equiv \frac{\mathrm{Var}(E_t r_{t+1} + \cdots + E_t r_{t+K})}{\mathrm{Var}(r_{t+1} + \cdots + r_{t+K})}. \tag{3.3}$$

Dividing by the one-period R^2 statistic and rearranging, one obtains

$$\frac{R^2(K)}{R^2(1)} = \left[\frac{\mathrm{Var}(E_t r_{t+1} + \cdots + E_t r_{t+K})}{\mathrm{Var}(E_t r_{t+1})}\right]\left[\frac{\mathrm{Var}(r_{t+1})}{\mathrm{Var}(r_{t+1} + \cdots + r_{t+K})}\right]. \tag{3.4}$$

The first ratio on the right hand side of Eq. (3.4) is just the square of the K-period regression coefficient divided by the square of the one-period regression coefficient. In the AR(1) example this is $(1-\phi^K)^2/(1-\phi)^2$, which is approximately equal to K^2 for large ϕ and small K. The second ratio on the right hand side of Eq. (3.4) depends on the autocorrelations of stock returns. Cochrane (1988) and Lo and MacKinlay (1988) have shown that

$$\frac{\mathrm{Var}(r_{t+1} + \cdots + r_{t+K})}{\mathrm{Var}(r_{t+1})} = K\left(1 + 2\sum_{j=1}^{K}\left(1 - \frac{j}{K}\right)\mathrm{Corr}(r_{t+1}, r_{t+1-j})\right). \tag{3.5}$$

The autocovariances calculated in Eq. (2.9) can be used to evaluate this expression. Unless σ_{dx} is large and positive, the autocorrelations of stock returns are all negative. It follows that the ratio in Eq. (3.5) is less than K so its reciprocal in Eq. (3.4) is greater than $1/K$.

Putting the two terms on the right hand side of Eq. (3.4) together, one finds that if expected stock returns are very persistent, the multi-period R^2 statistic grows at first approximately in proportion to the horizon K. Intuitively, this occurs because forecasts of expected returns several periods ahead are only slightly less variable than the forecast of next period's expected return, and they are perfectly correlated with it. Successive realized returns, on the other hand, are slightly negatively correlated with one another. Thus at first the variance of the multi-period fitted value grows more rapidly than the variance of the multi-period realized return, increasing the multi-period R^2 statistic. Eventually, of course, forecasts of returns in the distant future die out so the first ratio on the right hand side of Eq. (3.4) converges to a fixed limit; but the variability of realized multi-period returns continues to increase, so the second ratio on the right hand side of Eq. (3.4) becomes proportional to $1/K$. Thus eventually multi-period R^2 statistics go to zero as the horizon increases.

The exact expression for the multi-period R^2 statistic is algebraically complicated, even when one uses the approximation $\rho \approx 1$ and considers only a two-period horizon. In this case tedious but straightforward calculations yield

$$\frac{R^2(2)}{R^2(1)} = \frac{(1+\phi)^2}{2}\left[\frac{(1-\phi)\left(\sigma_{\mathrm{d}}^2/\sigma_x^2\right) + 2(1-\beta_{\mathrm{d}x})}{(1-\phi)\left(\sigma_{\mathrm{d}}^2/\sigma_x^2\right) + (1+\phi)(1-\beta_{\mathrm{d}x})}\right]. \tag{3.6}$$

The ratio in Eq. (3.6) approaches $(1+\phi)$ as $\sigma_{\mathrm{d}}^2/\sigma_x^2$ (and therefore $\beta_{\mathrm{d}x}$) approach zero, so a two-period regression may have an R^2 statistic almost twice that of a one-period regression if expected returns are persistent and highly variable. On the other hand, the ratio approaches $(1+\phi)^2/2$ as $\sigma_{\mathrm{d}}^2/\sigma_x^2$ approaches infinity, so a two-period regression may have an R^2 statistic only half that of a one-period regression if expected returns have only small and transitory variation.

Table 2 gives exact numerical results for horizons beyond two periods. The table is organized in exactly the same way as Table 1. For each set of parameter values, the table reports the horizon that maximizes the R^2 statistic, and the ratio of the maximized R^2 statistic to the R^2 of a one-period regression. The different

Table 2
Choosing horizon to maximize R^2, $K^* = \arg\max_K R^2(K)$, $(R^2(K^*)/R^2(1))$

A: $\beta_{\mathrm{d}x} = 0$

$R^2(1)/R^2_{\max}(1)$								
ϕ	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{\max}(1)$
0.100	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.02)	2 (1.06)	5 (1.11)	0.453
0.250	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.09)	2 (1.18)	3 (1.23)	8 (1.33)	0.378
0.500	2 (1.15)	2 (1.20)	3 (1.29)	3 (1.48)	4 (1.68)	5 (1.80)	14 (2.00)	0.254
0.750	5 (1.94)	5 (2.06)	6 (2.33)	7 (2.76)	10 (3.24)	12 (3.50)	28 (3.99)	0.129
0.900	12 (4.47)	13 (4.79)	15 (5.49)	20 (6.65)	26 (7.92)	31 (8.62)	58 (9.86)	0.055
0.950	25 (8.72)	27 (9.36)	32 (10.8)	40 (13.1)	52 (15.6)	61 (16.9)	92 (19.1)	0.030
0.980	65 (21.4)	69 (22.9)	78 (26.2)	96 (31.3)	117 (36.3)	130 (38.8)	152 (42.0)	0.016

B: Zero-autocorrelation $\beta_{\mathrm{d}x}$

$R^2(1)/R^2_{\max}(1)$								
ϕ ($\beta_{\mathrm{d}x}$)	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{\max}(1)$
0.100 (0.994)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	0.990
0.250 (0.992)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	0.937
0.500 (0.985)	2 (1.13)	2 (1.13)	2 (1.13)	2 (1.13)	2 (1.13)	2 (1.13)	2 (1.13)	0.750
0.750 (0.966)	4 (1.87)	4 (1.87)	4 (1.87)	4 (1.87)	4 (1.87)	4 (1.87)	4 (1.87)	0.437
0.900 (0.909)	12 (4.29)	12 (4.29)	12 (4.29)	12 (4.29)	12 (4.29)	12 (4.29)	12 (4.29)	0.190
0.950 (0.822)	24 (8.35)	24 (8.35)	24 (8.35)	24 (8.35)	24 (8.35)	24 (8.35)	24 (8.35)	0.098
0.980 (0.602)	62 (20.6)	62 (20.6)	62 (20.6)	62 (20.6)	62 (20.6)	62 (20.6)	62 (20.6)	0.040

rows in Table 2 correspond to different values of the persistence parameter ϕ. As one moves down the table, ϕ increases and both the R^2-maximizing horizon and the maximized R^2 increase. The different columns in Table 2 correspond to different degrees of predictability in one-period stock returns, relative to the maximum predictability consistent with each value of ϕ. As one moves to the right, stock returns become more predictable and again the R^2-maximizing horizon and maximized R^2 increase. The extreme case in panel A has $\phi = 0.98$ and no dividend uncertainty; in this case the R^2 statistic of a one-period regression is only 1.5% but the maximum R^2 statistic is $42 \times 1.5\% = 63\%$, obtained with a 152-period regression.

Turning to the weighted infinite-horizon regression, tedious but straightforward algebra shows that when $\rho = 1$ the R^2 statistic is

$$R^2(\gamma) = \left[\frac{(1-\gamma\phi)^2}{1-\gamma^2} \left(\frac{\sigma_{\mathrm{d}}^2}{\sigma_x^2} + 2\left[\frac{1}{1-\phi} - \frac{\gamma}{1-\gamma\phi} \right](1-\beta_{\mathrm{d}x}) \right) \right]^{-1}. \quad (3.7)$$

Note that R^2 is now written as a function of the weighting parameter γ rather than of the horizon K; hence the one-period regression R^2 is written as $R^2(0)$ because this is the case where $\gamma = 0$. Table 3 gives the value of γ that maximizes R^2, and the maximized ratio $R^2(\gamma)/R^2(0)$, for $\rho = 0.995$ and the same parameter values used before. As before, panel A of the table sets $\beta_{\mathrm{d}x} = 0$. When expected returns are highly variable, at the right of the table, the highest R^2 is obtained by setting γ very close to ρ in the manner of a volatility test. Note that when the one-period R^2 statistic is at its maximum in the right-hand column, the appropriately weighted infinite-horizon regression has an R^2 statistic of unity. This is because in the right-hand column the stock is like a consol whose long-run return is known with certainty.

3.2. Asymptotic statistical inference

Tables 2 and 3 show that increases in horizon can dramatically increase regression R^2 statistics when expected returns are persistent. Despite this appealing property, long-horizon regressions cannot be used without confronting some tricky statistical issues. The basic problem with a long-horizon regression is that when one uses all the data, the overlap in the dependent variable creates serial correlation in the residuals. If standard errors and test statistics are calculated without taking account of this, they greatly understate the true uncertainty about the regression coefficients.

Asymptotic statistical inference with serially correlated errors is by now well understood. The asymptotic theory is a special case of Hansen's (1982) Generalized Method of Moments, and was first applied to regressions with overlapping residuals by Hansen and Hodrick (1980). Hansen and Hodrick did not allow for

Table 3
Choosing discount factor to maximize R^2, $\gamma^* = \arg\max_\gamma R^2(\gamma)$, $(R^2(\gamma^*)/R^2(0))$

A: $\beta_{dx} = 0$

$R^2(0)/R^2_{max}(0)$								
ϕ	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(0)$
0.100	0.148 (1.02)	0.218 (1.05)	0.342 (1.13)	0.507 (1.28)	0.671 (1.51)	0.759 (1.67)	0.994 (2.21)	0.453
0.250	0.285 (1.09)	0.347 (1.14)	0.461 (1.25)	0.594 (1.46)	0.737 (1.76)	0.804 (1.96)	0.994 (2.64)	0.378
0.500	0.525 (1.38)	0.568 (1.47)	0.641 (1.67)	0.738 (2.02)	0.834 (2.51)	0.875 (2.84)	0.995 (3.84)	0.254
0.750	0.763 (2.39)	0.781 (2.58)	0.819 (3.01)	0.868 (3.75)	0.917 (4.81)	0.941 (5.49)	0.995 (7.73)	0.129
0.900	0.903 (5.53)	0.914 (6.00)	0.931 (7.11)	0.949 (8.98)	0.966 (11.7)	0.977 (13.4)	0.995 (18.2)	0.055
0.950	0.952 (10.8)	0.959 (11.7)	0.966 (13.9)	0.975 (17.6)	0.983 (22.8)	0.988 (26.0)	0.995 (33.2)	0.030
0.980	0.982 (26.5)	0.982 (28.7)	0.986 (33.7)	0.989 (41.7)	0.992 (51.6)	0.994 (56.7)	0.995 (64.2)	0.016

B: Zero-autocorrelation β_{dx}

$R^2(0)/R^2_{max}(0)$								
ϕ (β_{dx})	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(0)$
0.100 (0.994)	0.102 (1.01)	0.102 (1.01)	0.102 (1.01)	0.102 (1.01)	0.102 (1.01)	0.102 (1.01)	0.102 (1.01)	0.990
0.250 (0.992)	0.253 (1.07)	0.253 (1.07)	0.253 (1.07)	0.253 (1.07)	0.253 (1.07)	0.253 (1.07)	0.253 (1.07)	0.937
0.500 (0.985)	0.498 (1.33)	0.498 (1.33)	0.498 (1.33)	0.498 (1.33)	0.498 (1.33)	0.498 (1.33)	0.498 (1.33)	0.750
0.750 (0.966)	0.750 (2.29)	0.750 (2.29)	0.750 (2.29)	0.750 (2.29)	0.750 (2.29)	0.750 (2.29)	0.750 (2.29)	0.437
0.900 (0.909)	0.902 (5.26)	0.902 (5.26)	0.902 (5.26)	0.902 (5.26)	0.902 (5.26)	0.902 (5.26)	0.901 (5.26)	0.190
0.950 (0.822)	0.948 (10.3)	0.948 (10.3)	0.948 (10.3)	0.948 (10.3)	0.948 (10.3)	0.948 (10.3)	0.948 (10.3)	0.098
0.980 (0.602)	0.980 (25.2)	0.980 (25.3)	0.980 (25.3)	0.980 (25.3)	0.980 (25.3)	0.980 (25.3)	0.980 (25.3)	0.040

heteroskedasticity in the manner of White (1984), but it is now standard to do so. The heteroskedasticity correction plays no role in the asymptotic analysis here, since the AR(1) example is homoskedastic.

To state Hansen's result in appropriate generality, consider a regression of the form

$$Y_t = X_t\theta + e_t, \tag{3.8}$$

where X_t and θ may be vectors and the error term e_t may have arbitrary serial correlation. We define $W_t = e_t X_t$, the product of the equation error and the vector of explanatory variables, and use the notation $\hat{\theta}$ for the OLS estimate of θ in Eq. (3.8). Then the asymptotic distribution of $\hat{\theta}$ is normal, with mean θ and variance–covariance matrix given by

$$T\,\mathrm{Var}(\hat{\theta} - \theta) = Z_0^{-1} S_0 Z_0^{-1}, \tag{3.9}$$

where

$$Z_0 \equiv E[X_t X_t'], \tag{3.10}$$

$$S_0 \equiv E\left[\sum_{j=-\infty}^{\infty} W_t W_{t-j}'\right]. \tag{3.11}$$

S_0 is the spectral density matrix of W_t evaluated at frequency zero, or equivalently the two-sided sum of all the autocovariances of W_t.

In practice, of course, S_0 and Z_0 are unknown. However, they can be estimated in the usual way by using sample moments in place of population moments. Z_0 can be estimated straightforwardly by $Z_T \equiv (1/T)\Sigma_{t=1}^{T} X_t X_t'$, but estimation of S_0 requires that one truncate the infinite two-sided sum at some point K to get an estimate $S_T \equiv C_T(0) + \Sigma_{j=1}^{K-1}[C_T(j) + C_T(j)']$, where $C_T(j) \equiv (1/T)\Sigma_{t=j+1}^{T} W_t W_{t-j}'$. These estimates deliver asymptotically correct statistical inference provided that the truncation point K increases (but not too rapidly) with the sample size.

In the case of the unweighted long-horizon regression Eq. (3.1), these results simplify because the regression has demeaned variables and a scalar regressor, and because the assumption that x_t captures all predictability in returns implies that the autocovariances of the equation error are zero beyond the regression horizon K. Thus if (K) is the OLS estimate of $\beta(K)$ in Eq. (3.1), and w_{t+K} is the scalar product of the explanatory variable and the equation error, $w_{t+K} \equiv x_t e_{t+K,K}$, we have asymptotically that under the null hypothesis,

$$T\,\mathrm{Var}\left(\hat{\beta}(K) - \beta(K)\right) = \frac{E\left[\sum_{j=-K+1}^{K-1} W_{t+k} W_{t+k-j}\right]}{\sigma_x^4}. \tag{3.12}$$

In the single-period regression case $K = 1$, Eq. (3.12) becomes $\mathrm{Var}(e_{t+1,1})/\mathrm{Var}(x_t)$, the standard expression for the variance of a regression coefficient estimate with a scalar zero-mean regressor and serially uncorrelated residuals. Eq. (3.12) can be used to construct asymptotic standard errors for K-period regression coefficients.

3.3. Approximate slope of long-horizon regressions

The asymptotic theory summarized above can also be used to compare the power of short- and long-horizon regressions to reject the hypothesis that stock returns are unforecastable. A simple and intuitive measure of asymptotic power is the "approximate slope" of a test statistic, due to Bahadur (1960) and analyzed in greater detail by Geweke (1981). Jegadeesh (1991) and Richardson and Smith (1991a,b) have recently applied this concept to study regressions of returns on lagged returns over various horizons.

Approximate slope is calculated assuming that a specific fixed alternative hypothesis is true, and it measures the rate at which the null hypothesis becomes easier to reject as the sample size increases. Formally, if one fixes the power of a test against a given alternative hypothesis, the significance level s of the test decreases with sample size T. Approximate slope is the limit of the ratio $-2\log(s)/T$ as T increases. Geweke (1981) shows that the ratio of approximate slopes for two different test statistics can be interpreted as the limit, as s declines, of the reciprocal of the ratio of the number of observations each statistic needs to reject the null hypothesis with a given probability. Thus if test statistic A has approximate slope twice that of test statistic B, then asymptotically statistic A will require only half as many observations as statistic B to reject the null hypothesis with given probability.

Geweke (1981) also shows that for any test statistic that has a χ^2 distribution under the null hypothesis, the approximate slope is the probability limit of the statistic under the alternative hypothesis, deflated by the sample size. In the present application, the approximate slope $c(K)$ of a K-period regression test that $\beta(K)$ is zero is just the square of the population value of $\beta(K)$ under the alternative, divided by T times the asymptotic variance of (K) as given in Eq. (3.12):

$$c(K) \equiv \frac{\beta(K)^2}{T\,\mathrm{Var}\left(\hat{\beta}(K) - \beta(K)\right)}. \tag{3.13}$$

Thus K-period regressions can have asymptotic power advantages, in the sense captured by approximate slope, if the squared regression coefficient increases with K faster than the asymptotic variance in Eq. (3.12).

Approximate slope calculations in the AR(1) example are straightforward but are even more complicated algebraically than the calculation of R^2 statistics. If

one uses the simplifying assumption $\rho = 1$, the approximate slope of a one-period regression is

$$c(1) = \frac{1}{\sigma_{\mathrm{d}}^2/\sigma_x^2 + (1 + \phi - 2\beta_{\mathrm{d}x})/(1 - \phi)}, \tag{3.14}$$

while the slope of an unweighted overlapping two-period regression, divided by the slope of a one-period regression, is

$$\frac{c(2)}{c(1)} = \frac{(1+\phi)^2 + (1-\phi^2)\left(\sigma_{\mathrm{d}}^2/\sigma_x^2\right) - 2(1+\phi)\beta_{\mathrm{d}x}}{(1+3\phi^2) + 2(1-\phi)\left(\sigma_{\mathrm{d}}^2/\sigma_x^2\right) - 2(1+\phi)\beta_{\mathrm{d}x}}. \tag{3.15}$$

The right hand side of Eq. (3.15) approaches $(1+\phi)^2/(1+3\phi^2)$ as $\sigma_{\mathrm{d}}^2/\sigma_x^2$ (and therefore $\beta_{\mathrm{d}x}$) approach zero. $(1+\phi)^2/(1+3\phi^2)$ slightly exceeds one when $0 < \phi < 1$, reaching a maximum of 1.33 when $\phi = 0.33$; thus a two-period regression may have a larger approximate slope than a one-period regression if expected returns are highly variable. On the other hand, the right hand side of Eq. (3.15) approaches $(1+\phi)/2$ as $\sigma_{\mathrm{d}}^2/\sigma_x^2$ approaches infinity, so a two-period regression may have an approximate slope only half that of a one-period regression if expected returns have only small and transitory variation. The importance

Table 4
Choosing horizon to maximize approximate slope, $K^* = \arg\max_K c(K)$, $(c(K^*)/c(1))$

A: $\beta_{\mathrm{d}x} = 0$

$R^2(1)/R^2_{\max}(1)$								
ϕ	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{\max}(1)$
0.100	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.07)	5 (1.20)	0.453
0.250	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.13)	2 (1.22)	8 (1.48)	0.378
0.500	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.06)	3 (1.23)	4 (1.38)	13 (1.84)	0.254
0.750	1 (1.00)	1 (1.00)	1 (1.00)	3 (1.06)	6 (1.26)	9 (1.45)	27 (2.08)	0.129
0.900	1 (1.00)	1 (1.00)	1 (1.00)	6 (1.06)	16 (1.29)	24 (1.53)	55 (2.33)	0.055
0.950	1 (1.00)	1 (1.00)	2 (1.00)	14 (1.07)	35 (1.37)	49 (1.67)	89 (2.68)	0.030
0.980	1 (1.00)	1 (1.00)	3 (1.00)	48 (1.15)	94 (1.67)	117 (2.19)	153 (3.71)	0.016

B: Zero-autocorrelation $\beta_{\mathrm{d}x}$

$R^2(1)/R^2_{\max}(1)$								
ϕ ($\beta_{\mathrm{d}x}$)	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{\max}(1)$
0.100 (0.994)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	0.990
0.250 (0.992)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	1 (1.00)	0.937
0.500 (0.985)	1 (1.00)	1 (1.00)	1 (1.00)	2 (1.11)	2 (1.56)	2 (1.97)	2 (3.00)	0.750
0.750 (0.966)	1 (1.00)	1 (1.00)	2 (1.06)	4 (1.43)	4 (2.40)	5 (3.29)	5 (6.08)	0.437
0.900 (0.909)	1 (1.00)	1 (1.00)	5 (1.08)	9 (1.61)	12 (2.90)	12 (4.20)	13 (8.04)	0.190
0.950 (0.822)	1 (1.00)	1 (1.00)	10 (1.09)	19 (1.66)	24 (3.05)	25 (4.45)	26 (8.56)	0.096
0.980 (0.602)	1 (1.00)	1 (1.00)	24 (1.09)	49 (1.69)	61 (3.13)	64 (4.58)	67 (8.88)	0.040

of persistence can also be seen by noting that the right hand side of Eq. (3.15) approaches $(1+\sigma_{\mathrm{d}}^2/\sigma_x^2-2\beta_{\mathrm{d}x})/(1+2\sigma_{\mathrm{d}}^2/\sigma_x^2-2\beta_{\mathrm{d}x})$, which is always less than one, as ϕ approaches zero. Thus, long-horizon regressions cannot have a greater approximate slope than short-horizon regressions when expected returns follow a white noise process.

Table 4 gives approximate slopes for overlapping regressions with horizons beyond two periods. For each set of parameter values the table reports the horizon that maximizes the approximate slope, and the ratio of the maximized approximate slope to the approximate slope of a one-period regression. Both the maximizing horizon and the maximized approximate slope ratio increase as one moves down and to the right. In the case where $\beta_{\mathrm{d}x}=0$, $\phi=0.98$, and there is no dividend uncertainty, the regression with the greatest approximate slope has a horizon of 153 periods and the approximate slope ratio is 3.7. Taken together, Eq. (3.15) and Table 4 show that when expected returns are highly variable and persistent the approximate slope, as a function of the horizon, rises only slowly but continues to rise until the horizon becomes very long. Table 4 also shows that the slope-maximizing horizon is generally slightly less than the R^2-maximizing horizon in Table 2, and the proportional increase in approximate slope from using a long horizon is much less than the proportional increase in the R^2 statistic.

Turning to the weighted infinite-horizon regression, the approximate slope is again tedious but straightforward to compute when $\rho=1$:

$$c^*(\gamma)=\left[\frac{(1-\gamma^2\phi^2)}{1-\gamma^2}\left(\frac{\sigma_{\mathrm{d}}^2}{\sigma_x^2}+\frac{(1-\gamma)^2(1+\phi)}{(1-\gamma\phi)^2(1-\phi)}-2\frac{(1-\gamma)\beta_{\mathrm{d}x}}{(1-\gamma\phi)(1-\phi)}\right)\right]^{-1}. \tag{3.16}$$

Table 5 gives the value of γ that maximizes c^*, and the maximized ratio $c^*(\gamma)/c^*(0)$, for $\rho=0.995$ and the same parameter values used before. In the right-hand column the maximized slope ratio is infinite because the weighted infinite-horizon regression has an R^2 of unity and an infinite approximate slope. The slope-maximizing value of γ is fairly close to the R^2-maximizing value except at the far left of the table. Comparing the approximate slopes in Table 5 with those in Table 4, it is clear that weighted regressions can have large approximate slope advantages over unweighted long-horizon regressions. This, together with the result that γ should be close to ρ when $\beta_{\mathrm{d}x}=0$, may help to explain the dramatic empirical results obtained with volatility tests.

The approximate slope advantages of overlapping long-horizon regressions are striking for several reasons. First, in the example studied here both short- and long-horizon regressions use the same forecasting variable x_t. Previous work, notably Richardson and Smith (1991a,b), has studied regressions of returns on lagged returns in which the regressor changes as the return horizon changes. It is

Table 5
Choosing discount factor to maximize approximate slope, $\gamma^* = \arg\max_\gamma c(\gamma)$, $(c(\gamma^*)/c(0))$

A: $\beta_{dx} = 0$

$R^2(0)/R^2_{max}(0)$								
ϕ	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(0)$
0.100	0.051 (1.00)	0.127 (1.02)	0.276 (1.10)	0.459 (1.35)	0.654 (2.05)	0.746 (2.89)	0.995 (.)	0.453
0.250	0.048 (1.00)	0.135 (1.02)	0.297 (1.10)	0.498 (1.34)	0.697 (2.01)	0.787 (2.82)	0.995 (.)	0.378
0.500	0.052 (1.00)	0.144 (1.01)	0.338 (1.08)	0.580 (1.31)	0.782 (1.93)	0.852 (2.69)	0.995 (.)	0.254
0.750	0.055 (1.00)	0.158 (1.01)	0.420 (1.06)	0.730 (1.25)	0.886 (1.85)	0.952 (2.61)	0.995 (.)	0.129
0.900	0.060 (1.00)	0.171 (1.00)	0.545 (1.03)	0.881 (1.22)	0.952 (1.89)	0.972 (2.77)	0.995 (.)	0.055
0.950	0.064 (1.00)	0.190 (1.00)	0.694 (1.08)	0.944 (1.61)	0.978 (2.90)	0.985 (4.20)	0.995 (.)	0.030
0.980	0.069 (1.00)	0.238 (1.00)	0.914 (1.03)	0.981 (1.38)	0.991 (2.66)	0.993 (4.71)	0.995 (.)	0.016

B: Zero-autocorrelation β_{dx}

$R^2(0)/R^2_{max}(0)$								
ϕ (β_{dx})	0.10	0.25	0.50	0.75	0.90	0.95	1.00	$R^2_{max}(0)$
0.100 (0.994)	0.012 (1.00)	0.028 (1.00)	0.050 (1.00)	0.072 (1.02)	0.091 (1.08)	0.095 (1.18)	0.100 (.)	0.990
0.250 (0.992)	0.022 (1.00)	0.066 (1.00)	0.129 (1.03)	0.192 (1.13)	0.227 (1.48)	0.239 (2.06)	0.250 (.)	0.937
0.500 (0.985)	0.052 (1.00)	0.142 (1.02)	0.284 (1.10)	0.408 (1.43)	0.467 (2.64)	0.485 (4.64)	0.500 (.)	0.750
0.750 (0.966)	0.084 (1.00)	0.247 (1.02)	0.530 (1.18)	0.677 (1.79)	0.727 (3.98)	0.739 (7.57)	0.750 (.)	0.437
0.900 (0.909)	0.110 (1.00)	0.381 (1.02)	0.782 (1.20)	0.862 (1.97)	0.890 (4.70)	0.895 (9.17)	0.900 (.)	0.190
0.950 (0.822)	0.114 (1.00)	0.468 (1.01)	0.887 (1.21)	0.933 (2.03)	0.945 (4.92)	0.947 (9.66)	0.950 (.)	0.098
0.980 (0.602)	1.123 (1.00)	0.575 (1.01)	0.954 (1.21)	0.973 (2.06)	0.978 (5.05)	0.979 (9.94)	0.980 (.)	0.040

easy to see that the return horizon can change the power of a regression when it changes the regressor, but there is no such effect here.

Second, the approximate slope advantages of long-horizon regressions do not depend on the existence of mean-reversion in realized asset returns. Long-horizon regressions can have greater approximate slope even when realized returns are white noise. To see this for the two-period case with $\rho = 1$, set $\beta_{dx} = 1$ to make realized returns white noise and simplify Eq. (3.15) to obtain

$$\frac{c(2)}{c(1)} = \frac{(1-\phi^2)\left(\sigma_d^2/\sigma_x^2 - 1\right)}{(3\phi^2 - 2\phi - 1) + 2(1-\phi)\left(\sigma_d^2/\sigma_x^2\right)}. \tag{3.17}$$

When σ_d^2/σ_x^2 is at its minimum of $1/(1-\phi^2)$, this expression reaches a maximum of $\phi^2(1+\phi)/((1-\phi^2)(1-3\phi)+2\phi^2)$, which exceeds one when ϕ lies between 0.366 and one.

More generally, panels B of Tables 2–5 repeat panels A setting β_{dx} to make realized returns white noise. With $\rho = 0.995$ rather than one, the required value of β_{dx} is not exactly one but it depends only on ϕ and is reported in parentheses under each value of ϕ. In panel B of Table 2, shorter horizons maximize the R^2 of unweighted regressions and the gain in explanatory power is less dramatic than in panel A. Panel B of Table 4 shows that shorter horizons maximize approximate slope for unweighted regressions than is the case in panel A; however, the gain in approximate slope can be much greater than before. When $\phi = 0.98$ and dividend uncertainty is at its minimum, for example, a 67-period regression can have an approximate slope almost nine times greater than a one-period regression.

In panel B of Table 3, the highest R^2 for a weighted regression is obtained when the weighting coefficient γ is very close to ϕ rather than ρ as in panel A.[5] As before, in the right hand column of the table the highest possible R^2 of a weighted long-horizon regression is actually unity. Panel B of Table 5 shows that in this case approximate slope is maximized in a weighted regression by setting $\gamma = \phi$. Just as in panel A, this gives an infinite approximate slope.

Although the approximate slope advantages of overlapping long-horizon regressions do not require mean reversion in returns, they do require that innovations in returns are negatively correlated with future returns. Eq. (2.8) writes the overall serial correlation of returns as the sum of the serial correlation of the expected return x_t and the correlation of the return innovation $v_t + 1$ with future returns. The former is positive when $\phi > 0$. The latter is negative unless dividend news and expected return innovations are strongly positively correlated, and it must be negative if the overall serial correlation is to be zero.

Long-horizon regressions only have approximate slope advantages when this correlation of return innovations with future returns is negative. Intuitively, a

[5] This result can be obtained analytically when $\rho = 1$. In this case $\beta_{dx} = 1$, so Eq. (3.7) simplifies considerably and it is straightforward to show that $R^2(\gamma)$ is maximized when $\gamma = \phi$.

regression coefficient is precisely estimated when the regression error has low variance. The error variance of a long-horizon regression is reduced if unexpectedly high returns early in the forecast period tend to be associated with lower returns later in the forecast period, and this is the source of the greater approximate slope of long-horizon regressions. More formally, Eq. (2.9) shows that return innovations are uncorrelated with future returns when $\beta_{dx} = 1 + \phi$. Eq. (2.10) shows that this requires $\sigma_d^2/\sigma_x^2 \geq (1+\phi)/(1-\phi)$. When these conditions are imposed on the general formula for the approximate slope ratio $c(2)/c(1)$, Eq. (3.15), the ratio can never be greater than one.[6] Hence, long-horizon regressions have no approximate slope advantages in this case.

3.4. Some other variations of long-horizon regressions

This subsection considers several other variations of the basic long-horizon regression. In the AR(1) example, it turns out that these variations offer little or no advantage over a short-horizon regression.

As a first variation, consider using only *non-overlapping data*, that is, sampling the data every K periods when the regression horizon is K. This procedure is very simple because the regression error is serially uncorrelated under the null hypothesis; for this reason it has been used by Fama and French (1988b, 1989) among others. With non-overlapping data the approximate slope ratio $c(K)/c(1)$ is just

$$\frac{c(K)}{c(1)} = \frac{\beta(K)^2}{K\,\mathrm{Var}(e_{t+K,K})/\mathrm{Var}(e_{t+1,1})}, \tag{3.18}$$

where the division by K reflects the fact that the K-period regression has only $1/K$ times as many observations.

Straightforward algebra shows that when $\rho = 1$ the slope of a non-overlapping two-period regression, divided by the slope of a one-period regression, is

$$\frac{c(2)}{c(1)} = \left(\frac{1+\phi}{2}\right)\frac{(1+\phi)^2 + (1-\phi^2)(\sigma_d^2/\sigma_x^2) - 2(1+\phi)\beta_{dx}}{(1+3\phi^2) + \phi(1-\phi)^2 + 2(1-\phi)(\sigma_d^2/\sigma_x^2) - 2(1+\phi)\beta_{dx}}. \tag{3.19}$$

This differs from the formula for the overlapping regression in two ways. The whole formula is multiplied by $(1+\phi)/2$, and there is an extra term $\phi(1-\phi)^2$ in the denominator. Both these differences reduce the approximate slope in Eq. (3.19), reflecting the loss in efficiency from dropping the overlapping data. Both differences disappear as ϕ approaches one, because there is less extra information in the overlapping data when the regressor is highly persistent.[7]

[6] To show this, note that when $\beta_{dx} = 1+\phi$, the approximate slope ratio is one when $\sigma_d^2/\sigma_x^2 = 2\phi/(1+\phi)$. But this is smaller than the lower bound on σ_d^2/σ_x^2 imposed by Eq. (2.10).

[7] Hansen and Hodrick (1980) originally made this point, which has been restated recently by Boudoukh and Richardson (1993).

Even though the overlapping data contain little extra information when ϕ is close to one, this information turns out to be critical in generating large approximate slope advantages for long-horizon regressions. To see this, first consider the case where $\beta_{dx} = 0$. Under the simplifying assumption that $\rho = 1$, as σ_d^2/σ_x^2 goes to zero the right hand side of Eq. (3.19) approaches its maximum value of $(1+\phi)^2/2(1+\phi^2)$, which is always less than one. Thus in this case a long-horizon regression can have an approximate slope advantage over a short-horizon regression only if it uses overlapping data. Now keep $\beta_{dx} = 0$ but set $\rho = 0.995$. Numerical calculations like those in Table 4 show that in this case a non-overlapping long-horizon regression can offer a gain in approximate slope, but it is extremely small. At the bottom right corner of Table 4 panel A, for example, an overlapping long-horizon regression with the slope-maximizing horizon of 153 periods increases approximate slope by a factor of 3.71. For the non-overlapping regression these numbers are 21 and 1.02, respectively.

Next consider the case where β_{dx} is set to eliminate serial correlation in realized returns. Under the simplifying assumption that $\rho = 1$, as σ_d^2/σ_x^2 approaches its minimum value of $1/(1-\phi^2)$, the right hand side of Eq. (3.19) approaches its maximum value of $(1+\phi)\phi^2/(4/(1+\phi)+2(1+\phi)(\phi^2-1))$, which exceeds one when ϕ lies between 0.521 and 1. Thus in this case a non-overlapping long-horizon regression can have an approximate slope advantage over a short-horizon regression. Numerical calculations setting $\rho = 0.995$ show that the gain in approximate slope is much reduced by using non-overlapping data. At the bottom right hand corner of Table 4, panel B, an overlapping long-horizon regression with the slope-maximizing horizon of 67 periods increases approximate slope by a factor of 8.88; for the non-overlapping regression theses numbers are 46 and 1.96, respectively.

These calculations show that contrary to Boudoukh and Richardson (1993), the case for running long-horizon regressions depends on using all the data rather than sampling the data to eliminate the serial correlation of the regression error term.

A second variation of the standard long-horizon regression is a *backward-filtered short-horizon regression*. This keeps the short-horizon return as the dependent variable but uses a backward moving average of x_t as the regressor:

$$r_{t+1} = \theta(K)(x_t + \cdots + x_{t-K}) + v_{t+1,K}. \tag{3.20}$$

The numerator of the regression coefficient $\theta(K)$ in Eq. (3.20) is the same as the numerator of the regression coefficient $\beta(K)$ in Eq. (3.1), because the covariance of x measured at one data and r measured at another date depends only on the difference between the two dates. Hence, $\theta(K) = 0$ in Eq. (3.20) if and only if $\beta(K) = 0$ in Eq. (3.1).

Cochrane (1991) seems to argue, on the basis of this fact, that long-horizon regressions have no advantages over backward-filtered regressions of the form (Eq. (3.20)). He writes: "Intuitively, the filtered right-hand-variable regression [Eq. (3.20)] has better power than the raw [short-horizon] regression because a

slow-moving right-hand variable can pick up a slow-moving component on the left-hand side. This increase in power is the same as the increase one obtains by aggregating the left-hand variable [in a long-horizon regression]" (p. 476, words in square brackets added for clarification). In the context of the AR(1) example, however, it is easy to see that the backward-filtered regression Eq. (3.20) can have no advantages, because it merely replaces the optimal forecasting variable x_t with an inferior forecasting variable that averages x_t and its own lags. Hence, the characteristics of long-horizon regressions are not the same as those of backward-filtered regressions. A backward-filtered regression should be used when there is reason to believe that a moving average of some observed variable is a better proxy for the expected one-period stock return than the observed variable itself; the analysis of approximate slope suggests that a long-horizon regression should be used when there is reason to believe that the return contains a variable and persistent predictable component.

A third variation, proposed by Hodrick (1992) and Bollerslev and Hodrick (1992), is to run a long-horizon regression but to *impose the null hypothesis* when calculating the asymptotic standard error. The basic long-horizon standard error calculation assumes only that the autocorrelations of the regression error are zero beyond lag $K-1$, which is true under both the null and the alternative. But the null hypothesis also implies that the regression error $e_{t+K,K} = r_{t+1} + \ldots + r_{t+K}$, and that these returns are serially uncorrelated. Thus the cross-product $w_{t+K} \equiv x_t e_{t+K,K} = x_t(r_{t+1} + \ldots + r_{t+K})$. Substituting into Eq. (3.12) and simplifying, one obtains an alternative asymptotic variance formula as

$$T\mathrm{Var}\left(\hat{\beta}(K) - \beta(K)\right) = \frac{\mathrm{Var}(r_{t+1})\mathrm{Var}(x_{t+1} + \cdots + x_{t+K})}{\sigma_x^4}. \qquad (3.21)$$

If one uses this formula in place of Eq. (3.12), the approximate slope of K-period regression relative to a one-period regression becomes

$$\frac{c(K)}{c(1)} = \frac{\beta(K)^2\mathrm{Var}(x_{t+1})}{\mathrm{Var}(x_{t+1} + \cdots + x_{t+K})}, \qquad (3.22)$$

which declines with the horizon K. When $K=2$, Eq. (3.22) implies $c(2)/c(1) = (1+\phi)/2 < 1$. Thus the Bollerslev and Hodrick (1992) and Hodrick (1992) estimator eliminates the approximate slope advantages of the basic long-horizon regression. If the alternative hypothesis is true, then uncertainty is less at long horizons than one would think from looking only at short horizons; long-horizon regressions have greater approximate slope only when one allows the data to reveal this fact.

This result may also be relevant for the results of Jegadeesh (1991) and Richardson and Smith (1991b). Both these papers study regressions of returns on lagged returns, and both calculate standard errors imposing the null hypothesis.

Richardson and Smith (1991b) require that the return horizon be the same for the regressor and the dependent variable, while Jegadeesh (1991) relaxes this restriction. Jegadeesh finds that approximate slope is maximized for a long-horizon regressor but a short-horizon dependent variable. This may be due to the fact that Jegadeesh uses standard errors which impose the null hypothesis.

4. Monte Carlo results

In this section I undertake a preliminary exploration of the finite-sample size and power properties of long-horizon regressions. While finite-sample performance is always the question of ultimate interest, and asymptotic theory is at best a guide to such performance, there are several reasons why in this case it is particularly important to explore finite-sample performance directly through Monte Carlo experiments.

First, it is well-known that the standard error correction of Hansen and Hodrick (1980) performs poorly when the degree of overlap is large relative to the sample size. Related standard error corrections, such as that of Newey and West (1987), also deteriorate in these circumstances.

Second, Stambaugh (1999) has analyzed one-period regressions of the type studied here and has quantified the finite-sample bias in their coefficient estimates. In the notation of this paper, Stambaugh's result is

$$E\left[\hat{\beta}(1)-\beta(1)\right]=-\left[\frac{1+3\phi}{T}\right]\frac{\sigma_{uv}}{\sigma_u^2}, \tag{4.1}$$

where u is the innovation in the expected return and v is the unexpected stock return. The term in square brackets is the standard formula for downward bias in an AR(1) coefficient estimate; Eq. (4.1) shows that the effect of this on the bias of $\hat{\beta}(1)$ depends on the covariance structure of u and v. Using Eq. (2.5) to break the unexpected stock return into its components, Eq. (4.1) can be rewritten as

$$E\left[\hat{\beta}(1)-\beta(1)\right]=-\left[\frac{1+3\phi}{T}\right]\left(\frac{\beta_{dx}}{1-\phi^2}-\frac{1}{1-\phi}\right). \tag{4.2}$$

When $\beta_{dx}=0$, this simplifies to $(1+3\phi)/(1-\phi)T$. It is easy to see that the finite-sample bias is increasing in ϕ; thus one-period regressions are most seriously biased in precisely those cases where long-horizon regressions have the greatest asymptotic advantages. This result is troubling even though it does not directly concern the properties of test statistics or long-horizon regressions.

A third reason to be particularly concerned about finite-sample behavior is that the analysis of approximate slope differs from standard asymptotic analysis.

Approximate slope is calculated for fixed alternatives, and long-horizon regressions only have increased approximate slope under alternatives that are sufficiently different from the null hypothesis. But these are precisely the alternatives under which it is easiest to reject the null hypothesis using a one-period regression. Thus in very large samples any gain in power may be unnecessary, since even a one-period regression can reject the null at any conventional significance level; power advantages in smaller samples are essential if there is to be a practical argument for using long-horizon regressions.

Another way to understand this point is to note that T times the R^2 statistic of a one-period regression is asymptotically distributed $\chi^2(1)$ under the null hypothesis that returns are unpredictable. Tables 2–5 give the probability limit of the R^2 statistic of a one-period regression for each alternative model; this should be roughly the median R^2 statistic that will be obtained in repeated finite samples under the alternative. We can compare T times this value with 3.8, say, the 5% critical value of a $\chi^2(1)$, to get a feel for the finite-sample power of a one-period regression. For example, consider the alternatives at the far right of each table, for which the one-period R^2 statistic is at its maximum. When $\phi = 0.5$ in panel A, $R^2(1) = 0.254$ and $TR^2(1) = 3.8$ for $T = 15$, suggesting that a one-period regression will reject at the 5% level half the time with 15 observations. When $\phi = 0.9$, $R^2(1) = 0.055$ and $TR^2(1) = 3.8$ for $T = 69$, so now almost 70 observations are needed for a one-period regression to reject with 50% probability. When $\phi = 0.95$ and 0.98, similar calculations show that 127 and 253 observations are needed. Only in these last two cases do one-period regressions have low power in samples of typical size, and in these cases long-horizon regressions must have very long horizons to get substantial power advantages.

Since the asymptotic critical values of long-horizon test statistics may be incorrect in finite samples, it is important to calculate empirical critical values under the null hypothesis before evaluating the ability of these statistics to reject the null when a particular alternative hypothesis is true. Unfortunately, this task is not straightforward because the hypothesis that returns are unpredictable is consistent with any time-series behavior for the regressor and any correlation between innovations in the regressor and innovations in returns. Empirical critical values are likely to be different for different unpredictable-return processes.

The Monte Carlo study reported here deals with this problem as follows. Starting with a process for x_t, the alternative hypothesis is that returns satisfy Eq. (2.5): $r_{t+1} = x_t + v_{t+1} = x_t + v_{\mathrm{d},t+1} - \rho u_{t+1}/(1 - \rho\phi)$. To generate empirical critical values, one-period returns are first generated from the unpredictable process.

$$r_{t+1} = v_{t+1} = v_{\mathrm{d},t+1} - \frac{\rho u_{t+1}}{1 - \rho\phi}. \tag{4.3}$$

This is equivalent to setting $\lambda = 0$ in the original return Eq. (2.1). These one-period returns are then cumulated to get multi-period returns. Short- and

Table 6
Monte Carlo simulations, unweighted regressions, $\phi = 0.95$, $R^2(1)/R^2_{max}(1) = 1$

A: $\beta_{dx} = 0$

		Empirical critical values			Size-adjusted rejection rates		
		0.5%	1%	5%	0.5%	1%	5%
$T = 100$	$K = 1$	3.8	3.1	2.4	0.02	0.10	0.38
	2	4.1	3.3	2.6	0.02	0.10	0.37
	5	5.2	4.0	2.9	0.02	0.09	0.33
	10	6.6	5.1	3.7	0.02	0.10	0.32
	20	7.7	6.2	4.6	0.04	0.13	0.36
$T = 200$	$K = 1$	3.4	2.9	2.2	0.17	0.46	0.88
	2	3.6	2.9	2.3	0.14	0.47	0.87
	5	3.8	3.2	2.4	0.15	0.41	0.84
	10	4.7	3.7	2.7	0.10	0.33	0.78
	20	5.6	4.5	3.2	0.16	0.36	0.73
	40	7.0	5.3	3.9	0.12	0.38	0.76

B: Zero-autocorrelation β_{dx}

		Empirical critical values			Size-adjusted rejection rates		
		0.5%	1%	5%	0.5%	1%	5%
$T = 50$	$K = 1$	4.0	3.2	2.5	0.01	0.04	0.18
	2	4.6	3.7	2.8	0.01	0.04	0.16
	5	6.4	5.1	3.7	0.01	0.04	0.15
	10	9.3	7.1	5.1	0.02	0.05	0.15
$T = 100$	$K = 1$	3.6	3.1	2.4	0.03	0.10	0.33
	2	3.8	3.3	2.5	0.03	0.10	0.32
	5	4.6	3.8	2.8	0.03	0.09	0.29
	10	5.9	4.8	3.5	0.03	0.09	0.28
	20	7.8	6.1	4.4	0.03	0.10	0.28

long-horizon returns are regressed on x_t, and unpredictability is tested. The empirical distribution of the test statistics gives empirical critical values.[8]

Table 6 reports the results of this procedure for two alternative processes. In panel A the alternative process has $\beta_{dx} = 0$, $\phi = 0.95$, and $R^2(1)$ equal to its maximum value of 0.030. Sample sizes of 100 and 200, and horizons of 1, 2, 5,

[8] This procedure closely matches the properties of single-period returns under the alternative hypothesis, but does not well match the properties of multi-period returns under the alternative hypothesis. A procedure that avoids this problem is to generate both single- and multi-period returns under the alternative hypothesis, then subtract the predictable components from these returns. I implemented this procedure in an earlier version of this paper. It gives smaller empirical critical values for long-horizon regressions, and hence stronger evidence for power advantages of long-horizon regressions, than the procedure described in the text. The difference between the results of the two procedures is particularly marked for unweighted long-horizon regressions.

10, 20, and 40, are considered. The horizon is arbitrarily restricted to be no greater than 20% of the sample size, and so a 40-period horizon is used only when $T = 200$. 10,000 replications are used in the Monte Carlo experiment. Panel B is similar except that the alternative sets $\beta_{dx} = 0.822$ to make realized returns white noise, $R^2(1)$ equals its maximum value of 0.098, and sample sizes of 50 and 100 are considered. All test statistics are calculated using Hansen–Hodrick standard errors unless these are negative (a very rare occurrence), in which case the correction suggested by Newey and West (1987) is used.

The first three columns of the table show the empirical critical values for 0.5%, 1%, and 5% tests of unpredictable stock returns. These critical values increase strongly with the horizon. The remaining columns of the table report rejection rates, based on empirical critical values. In some parts of the table there is a very slight tendency for the rejection rates to increase with the horizon, but overall the table does not make a good case for power advantages of long-horizon regressions.[9]

Table 7 repeats Table 6 for weighted long-horizon regressions. Here the dependent variable regressed on x_t is always a weighted sum of returns from t to the end of the sample. The weight for return r_{t+1+i} is γ^i. Standard errors are corrected using the Hansen–Hodrick correction with K^* autocovariances, where K^* is either $T/5$ or the smallest number such that $\gamma^{K^*} \leq 0.05$, whichever is smaller. This procedure for choosing the number of autocovariances is both consistent and roughly in line with empirical practice. As in Table 6, Newey–West standard errors are used only if Hansen–Hodrick standard errors are negative, something which occurs very rarely.

Table 7 shows that weighted long-horizon regressions have striking power advantages even after correcting for finite-sample size distortion. The most dramatic results are found in panel A where $\beta_{dx} = 0$ and returns are mean-reverting. Here a one-period regression rejects the null only 3% of the time at the 0.5% level, while a weighted regression with weight 0.995 rejects 87% of the time. In panel B there is a less dramatic but still marked increase in power with the horizon of the regression. Interestingly, in this panel there is a detectable peak in power around $\gamma = 0.95$; this is what the approximate slope calculation in Table 5 would imply.

The results in Table 7 suggest that weighted long-horizon regressions may have important practical advantages in detecting predictable components of asset returns. However, these results rely on empirical critical values that were calculated knowing the true value of the persistence parameter ϕ. In practice the econometrician does not know ϕ but must estimate it. If the empirical critical values for

[9] I also studied the finite-sample behavior of non-overlapping long-horizon regressions, but to save space these results are not reported. The rejection rates of non-overlapping regressions fall with the horizon, slowly at first and then sharply when the horizon reaches 10% or 20% of the sample size.

Table 7
Monte Carlo simulations, weighted regressions, $\phi = 0.95$, $R^2(0)/R^2_{max}(0) = 1$

A: $\beta_{dx} = 0$

		Empirical critical values			Size-adjusted rejection rates		
		0.5%	1%	5%	0.5%	1%	5%
$T = 100$	$\gamma = 0.00$	3.7	3.1	2.4	0.03	0.12	0.39
	0.75	5.7	4.6	3.3	0.03	0.12	0.37
	0.90	8.2	6.4	4.4	0.05	0.14	0.40
	0.95	11.0	8.2	5.5	0.13	0.26	0.56
	0.98	14.2	10.4	7.1	0.39	0.57	0.77
	0.995	17.0	12.5	8.7	0.87	0.96	0.99
$T = 200$	0.00	3.5	2.9	2.2	0.13	0.46	0.87
	0.75	4.6	3.7	2.8	0.15	0.47	0.87
	0.90	5.9	4.5	3.2	0.18	0.43	0.81
	0.95	7.0	5.5	4.0	0.27	0.53	0.83
	0.98	9.5	7.2	4.9	0.57	0.76	0.90
	0.995	10.6	8.4	6.0	0.90	0.97	0.99

B: zero-autocorrelation β_{dx}

		Empirical critical values			Size-adjusted rejection rates		
		0.5%	1%	5%	0.5%	1%	5%
$T = 50$	$\gamma = 0.00$	3.8	3.2	2.5	0.01	0.04	0.15
	0.75	7.1	5.8	4.1	0.01	0.05	0.16
	0.90	10.3	8.1	5.9	0.02	0.07	0.21
	0.95	13.3	10.7	7.6	0.03	0.09	0.26
	0.98	18.2	14.2	9.7	0.03	0.10	0.26
	0.995	22.7	16.6	11.6	0.02	0.08	0.20
$T = 100$	0.00	3.7	3.1	2.4	0.02	0.09	0.35
	0.75	5.7	4.6	3.3	0.03	0.09	0.35
	0.90	8.2	6.4	4.4	0.04	0.12	0.37
	0.95	11.0	8.2	5.5	0.07	0.18	0.41
	0.98	14.2	10.4	7.1	0.09	0.23	0.45
	0.995	17.0	12.5	8.7	0.07	0.17	0.36

long-horizon regressions are more sensitive to ϕ than are the empirical critical values for short-horizon regression, then the need to estimate ϕ may be a serious difficulty in implementing long-horizon regression tests.[10]

Table 8 confirms that indeed the empirical critical values for regression tests do increase more rapidly with ϕ when the horizon is long (that is, when the discount factor γ is close to one). The table reports the 5% empirical critical value of a long-horizon test statistic for the standard set of γ values and ϕ ranging from 0.90 to 0.999. Variation in ϕ over this range raises the 5% critical value of a

[10] I am grateful to Jim Stock for making this point.

Table 8
Sensitivity of empirical critical values for weighted regressions, to the persistence of the forecasting variable

5% empirical critical value						
		ϕ				
		0.90	0.95	0.98	0.99	0.999
$T = 100$	$\gamma = 0.00$	2.3	2.4	2.7	2.8	2.9
	0.75	3.0	3.3	3.7	3.9	4.2
	0.90	3.9	4.4	5.0	5.5	6.1
	0.95	4.6	5.5	6.8	7.9	9.4
	0.98	5.4	7.1	10.1	13.0	19.0
	0.995	6.2	8.7	15.0	23.3	58.0
$T = 200$	0.00	2.2	2.2	2.5	2.6	2.9
	0.75	2.6	2.8	3.1	3.3	3.7
	0.90	3.1	3.2	3.6	3.9	4.5
	0.95	3.5	4.0	4.6	5.1	6.2
	0.98	4.2	4.9	6.3	7.6	10.9
	0.995	4.8	6.0	9.4	13.1	30.1

short-horizon ($\gamma = 0$) test from 2.3 to 2.9, but it raises the critical value of a long-horizon ($\gamma = 0.995$) test from 6.2 to 58!

This sensitivity of critical values to persistence certainly complicates the implementation of long-horizon regression tests, but it does not altogether eliminate their advantages. In practice an econometrician will use a bootstrap procedure, first estimating ϕ, then calculating critical values appropriate for that ϕ, and finally running the long-horizon regression test (Mark (1995) is a recent example of this approach). A Monte Carlo study, reported in Table 9, shows that this bootstrap procedure can have good finite-sample properties.

The simulated procedure first estimates ϕ by OLS regression of x_t on its own first lag. In panel A the estimated ϕ is used without any correction, while in panel B the estimate is adjusted upwards by $(1 + 3)/T$, the standard formula for downward bias in an AR(1) parameter. An empirical critical value appropriate for the estimated ϕ is then calculated. In practice this would be done by a Monte Carlo study, but it is computationally too burdensome to simulate the properties of this. Instead I first calculated Monte Carlo critical values for $\phi = 0.8$, 0.9, 0.95, 0.98, 0.98, and 0.999,and then used linear interpolation for intermediate values of ϕ.

The left hand side of Table 9 reports the rejection rates of this procedure when the null hypothesis is true, that is when returns are generated by Eq. (4.3) and have no predictable component. The right hand side reports the rejection rates under the alternative hypothesis that returns are predictable, with one-period R^2 statistic of 3% and perfect negative correlation between the innovations u_{t+1} and v_{t+1}. The

Table 9
Empirical size and power of a weighted regression bootstrap procedure, $T = 100$

A: No bias adjustment

		Rejection rate under H_0			Rejection rate under H_A		
		0.5%	1%	5%	0.5%	1%	5%
$\phi = 0.95$	$\gamma = 0.00$	0.00	0.02	0.07	0.03	0.14	0.41
	0.95	0.01	0.02	0.08	0.16	0.29	0.56
	0.98	0.01	0.03	0.10	0.42	0.58	0.76
	0.995	0.02	0.05	0.11	0.84	0.90	0.96
$\phi = 0.999$	0.00	0.00	0.01	0.07	0.01	0.03	0.11
	0.95	0.01	0.03	0.09	0.02	0.06	0.16
	0.98	0.02	0.05	0.14	0.07	0.14	0.26
	0.995	0.11	0.15	0.23	0.14	0.17	0.21

B: Bias adjustment

		Rejection rate under H_0			Rejection rate under H_A		
		0.5%	1%	5%	0.5%	1%	5%
$\phi = 0.95$	$\gamma = 0.00$	0.00	0.01	0.05	0.03	0.11	0.31
	0.95	0.01	0.02	0.06	0.10	0.19	0.39
	0.98	0.01	0.02	0.06	0.24	0.34	0.51
	0.995	0.01	0.03	0.07	0.56	0.63	0.71
$\phi = 0.999$	0.00	0.00	0.01	0.05	0.01	0.03	0.09
	0.95	0.00	0.01	0.06	0.01	0.03	0.11
	0.98	0.01	0.02	0.07	0.02	0.05	0.15
	0.995	0.02	0.03	0.07	0.04	0.07	0.10

properties of short- and long-horizon regressions are compared for two values of ϕ, 0.95 and 0.999. Throughout the table the sample size T is 100, and 10,000 replications are used.

The left hand side of Panel A of Table 9 shows modest size distortions for the unadjusted bootstrap procedure when the true value of ϕ is 0.95, and more serious distortions when the true value of ϕ is 0.999. The distortions worsen with the regression horizon. These results are not surprising, as the downward bias in estimated ϕ will have more serious effects when the true value of ϕ is very close to one. Fortunately, the simple bias correction in panel B seems to eliminate most of the size distortions in the bootstrap approach.

The right hand side of Table 9 shows that long-horizon regression tests continue to offer large power advantages even when bootstrapped critical values are used. Focusing on panel B, a test with $\gamma = 0.995$ rejects the null at the 0.5% level 56% of the time whereas a standard short-horizon test rejects at this level only 3% of the time. The long-horizon 0.5% test has a higher rejection rate even than a 1% or 5% short-horizon test.

5. Conclusion

In this paper I have tried to fill a surprising gap in the vast literature on long-horizon regressions. These regressions have become enormously popular with financial econometricians because of the striking results that they often generate. Despite this, there has been almost no theoretical work on the power properties of long-horizon regressions, so empirical work has proceeded without adequate motivation.[11]

I have used a simple AR(1) example to study the behavior of long-horizon regressions. When expected returns are both variable and highly persistent, a short-horizon regression will have a very low R^2 statistic but much higher R^2 statistics can be obtained at longer horizons. Under the same conditions approximate slope, a measure of asymptotic power, is increasing with the horizon over some interval. It is noteworthy that these results do not depend on mean reversion (negative serial correlation in realized returns), although they do depend on negative correlation between innovations in the regressor and in returns.

The paper also explores weighted long-horizon regressions, in which more distant future returns are downweighted relative to near future returns. These regressions are closely related to volatility tests, and they turn out to have even greater approximate slopes than the standard unweighted long-horizon regressions.

Long-horizon standard error corrections are known to have serious size distortions in finite samples. In addition, the circumstances under which long-horizon regressions have approximate slope advantages are also circumstances under which short-horizon regression coefficients have serious finite-sample bias. Accordingly I use Monte Carlo methods to see whether power advantages are available in finite samples. I first assume that the true persistence of the regressor is known. In this case unweighted long-horizon regressions do not seem to offer any significant power gains over short-horizon regressions, but weighted long-horizon regressions perform extremely well. I next assume that the econometrician does not know the persistence of the regressor but must estimate it. In this situation a bootstrap approach works well if one corrects for the downward finite-sample bias in the OLS estimate of the persistence parameter. Once again weighted long-horizon regressions offer striking power advantages.

The analysis of the paper can be extended in several directions. First, it would be desirable to link the approximate slope analysis used here to more conventional asymptotic theory, perhaps by using the concept of "point-optimal" tests. Second, one might relax the assumption used here that returns are homoskedastic. Stam-

[11] Jegadeesh (1991) and Richardson and Smith (1991a,b) have studied the approximate slopes of regressions in which long-horizon returns are regressed on lagged long-horizon returns, but these are special in that both the regressor and the dependent variable change with the horizon. In addition, these authors compute standard errors under the null rather than the alternative.

baugh (1993) finds that conditional heteroskedasticity increases the advantages of long-horizon regressions. Third, it would be interesting to study the properties of long-horizon statistics calculated from estimates of a VAR system. Campbell and Shiller (1988a,b) and Campbell (1991) find that such statistics reject very strongly the hypothesis that stock returns are unpredictable, and they report some Monte Carlo evidence that this is not due merely to size distortions, but the power properties of the statistics remain unclear.

References

Bahadur, R.R., 1960. Stochastic comparison of tests. Annals of Mathematical Statistics 31, 276–294.

Bollerslev, T., Hodrick, R.J., 1995. Financial market efficiency tests. In: Pesaran, M., Hashem, Wickens, Michael, R. (Eds.), Handbook of Applied Econometrics, vol. 1. Blackwell, Oxford.

Boudoukh, J., Richardson, M., The Statistics of Long-Horizon Regressions Revisited, unpublished paper, New York University and University of Pennsylvania, 1993.

Campbell, J.Y., 1991. A variance decomposition for stock returns. Economic Journal 101, 157–179.

Campbell, J.Y., Shiller, R.J., 1988a. The dividend–price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195–228.

Campbell, J.Y., Shiller, R.J., 1988b. Stock prices, earnings, and expected dividends. Journal of Finance 43, 661–676.

Campbell, J.Y., Shiller, R.J., 1991. Yield spreads and interest rate movements: a bird's eye view. Review of Economic Studies 58, 495–514.

Campbell, J.Y., Lo, A.W., MacKinlay, A., 1997. Present value relations. Chapter 7 in The Econometrics of Financial Markets. Princeton University Press, Princeton.

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

Cochrane, J.H., 1991. Volatility tests and efficient markets: a review essay. Journal of Monetary Economics 27, 463–484.

Engle, R.F., 1984. Wald, likelihood ratio and lagrange multiplier tests in econometrics. In: Griliches, Z., Intriligator, M.D. (Eds.), Chapter 13: Handbook of Econometrics, vol. II. North Holland, Amsterdam.

Fama, E.F., 1990. Term structure forecasts of interest rates, inflation, and real returns. Journal of Monetary Economics 25, 59–76.

Fama, E.F., Bliss, R., 1987. The information in long-maturity forward rates. American Economic Review 77, 680–692.

Fama, E.F., French, K.R., 1988a. Permanent and temporary components of stock prices. Journal of Political Economy 96, 246–273.

Fama, E.F., French, K.R., 1988b. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–24.

Fama, E.F., French, K.R., 1989. Business conditions and expected returns on stocks and bonds. Journal of Financial Economics 25, 23–49.

Faust, J., 1992. When are variance ratio tests for serial dependence optimal? Econometrica 60, 1215–1226.

Geweke, J., 1981. The approximate slope of econometric tests. Econometrica 49, 1427–1442.

Goetzmann, W.N., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hansen, L.P., Hodrick, R.J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. Journal of Political Economy 88, 829–853.
Hodrick, R.J., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.
Jegadeesh, N., 1991. Seasonality in stock price mean reversion: evidence from the U.S. and the U.K. Journal of Finance 46, 1427–1444.
LeRoy, S.F., Porter, R., 1981. The present value relation: tests based on variance bounds. Econometrica 49, 555–574.
Lo, A.W., MacKinlay, A., 1988. Stock prices do not follow random walks: evidence from a simple specification test. Review of Financial Studies 1, 41–66.
Lo, A.W., MacKinlay, A., 1989. The size and power of the variance radio test in finite samples: a Monte Carlo investigation. Journal of Econometrics 40, 203–238.
Mankiw, N.G., Shapiro, M.D., 1986. Do we reject too often? Economics Letters 20, 139–145.
Mark, N.C., 1995. Exchange rates and fundamentals: evidence on long-horizon predictability. American Economic Review 85, 201–218.
Mishkin, F.S., 1990a. What does the term structure tell us about future inflation? Journal of Monetary Economics 25, 77–95.
Mishkin, F.S., 1990b. The information in the longer maturity term structure about future inflation. Quarterly Journal of Economics 105, 815–828.
Nelson, C.R., Myung, J.K., 1993. Predictable stock returns: the role of small sample bias. Journal of Finance 48, 641–661.
Newey, W.K., West, K.D., 1987. A simple, positive definite, heteroscedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Poterba, J., Summers, L.H., 1988. Mean reversion in stock returns: evidence and implications. Journal of Financial Economics 22, 27–60.
Richardson, M., Smith, T., 1991a. Robust power calculations with tests for serial correlation in stock returns, unpublished paper, University of Pennsylvania and Duke University.
Richardson, M., Smith, T., 1991b. Tests of financial models in the presence of overlapping observations. Review of Financial Studies 4, 227–254.
Richardson, M., Stock, J.H., 1989. Drawing inferences from statistics based on multi-year asset returns. Journal of Financial Economics 25, 323–348.
Scott, L.O., 1985. The present value model of stock prices: regression tests and Monte Carlo results. Review of Economics and Statistics 67, 599–604.
Shiller, R.J., 1981. Do stock prices move too much to be justified by subsequent changes in dividends? American Economic Review 71, 421–436.
Shiller, R.J., 1989. Market Volatility. MIT Press, Cambridge, MA.
Stambaugh, R.F., 1993. Estimating conditional expectations when volatility fluctuates, National Bureau of Economic Research Technical Paper 140.
Stambaugh, R.F., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
Summers, L.H., 1986. Does the stock market rationally reflect fundamental values? Journal of Finance 41, 591–601.
White, H., 1984. Asymptotic Theory for Econometricians. Academic Press, Orlando, FL.