Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Financial Economics 68 (2003) 201–232
JOURNAL OF Financial ECONOMICS
www.elsevier.com/locate/econbase

Long-horizon regressions: theoretical results and applications ☆

Rossen Valkanov

The Anderson School of Management, University of California, 110 Westood Plaza, C5.19, Los Angeles, CA 90095, USA

Received 14 November 1999; received in revised form 1 February 2001

Abstract

I use asymptotic arguments to show that the t-statistics in long-horizon regressions do not converge to well-defined distributions. In some cases, moreover, the ordinary least squares estimator is not consistent and the R^2 is an inadequate measure of the goodness of fit. These findings can partially explain the tendency of long-horizon regressions to find "significant" results where previous short-term approaches find none. I propose a rescaled t-statistic, whose asymptotic distribution is easy to simulate, and revisit some of the long-horizon evidence on return predictability and of the Fisher effect.

JEL classification: C12; C22; G12; E47

Keywords: Predictive regressions; Long-horizon predictions; Stock returns; Fisher effect

1. Introduction

There has been an increasing interest in long-horizon regressions, because studies using long-horizon variables seem to find significant results where previous "short-term" approaches have failed. For example, Fama and French (1988), Campbell and Shiller (1987, 1988), Mishkin (1990, 1992), Boudoukh and Richardson (1993), and

☆ I have benefited from detailed comments by Mark Watson, Yacine Ait-Sahalia, Sergei Sarkissian, and an anonymous referee, and from discussions with Maria Gonzalez, Pedro Santa-Clara, Walter Torous, Ivo Welch, and participants at the 2000 Western Finance Association meetings in Sun Valley, Idaho. All remaining errors are my own.

E-mail address: rossen.valkanov@anderson.ucla.edu (R. Valkanov).

0304-405X/03/$ - see front matter
doi:10.1016/S0304-405X(03)00065-5

Fisher and Seater (1993), all studies with long-run variables, have received a lot of attention in finance and economics. The results in those papers are based on long-horizon variables, where the long-horizon variable is a rolling sum of the original series. In the literature, it is heuristically argued that long-run regressions produce more accurate results by strengthening the signal coming from the data while eliminating the noise. Whether the focus is on excess returns/dividend yields, the Fisher effect, or neutrality of money, the striking results produced by such studies prompted me to scrutinize the appropriateness of the econometric methods.

In this paper, I show that long-horizon regressions will always produce "significant" results, whether or not there is a structural relation between the underlying variables. To understand this conclusion, notice that in a rolling summation of series integrated of order zero (or $I(0)$), the new long-horizon variable behaves asymptotically as a series integrated of order one (or $I(1)$). Such persistent stochastic behavior will be observed whenever the regressor, the regressand, or both are obtained by summing over a nontrivial fraction of the sample. Based on this insight, I use the Functional Central Limit Theorem (FCLT) to analyze the distributions of statistics from long-run regressions commonly used in economics and finance. I find that, in addition to incorrect testing, overlapping sums of the original series might lead to inconsistent estimators and to a coefficient of determination, R^2, that does not converge to one in probability. These results are reminiscent of, but not analogous to, the result in Granger and Newbold (1974) and Phillips (1986, 1991) and those recently discussed by Ferson et al. (2003) in forecasting excess returns. The analogy lies in finding a spurious correlation between persistent variables when they are in fact statistically independent. However, there are two major differences. First, in long-horizon regressions, the rolling summation alters the stochastic order of the variables, resulting in unorthodox limiting distributions of the slope estimator, its t-statistic, and the R^2. More importantly, even if there is an underlying relationship between variables, the t-statistic will tend to reject it. In other words, estimation and testing using long-horizon variables cannot be carried out using the usual regression methods.

I provide a simple guide on how to conduct estimation and inference using long-horizon regressions. Based on previous empirical studies, I classify such regressions into four cases. The proposed classification emerges naturally from a consideration of null hypotheses and the persistence of the regressors. It allows a systematic analysis of the small-sample properties of long-horizon regressions. Those properties are analyzed by using the FCLT to derive accurate approximations of the small-sample distributions of the OLS estimator of the slope coefficient, its t-statistic and the coefficient of determination R^2. The estimators from some regressions, frequently used in empirical work, are not consistent. Moreover, the t-statistics from all considered regressions do not converge to well-defined distributions, thus calling into question the conclusions from studies that use long-run variables. The analytical results yield exact rates of convergence or divergence and permit us to modify the statistics in order to conduct correctly sized tests.

I propose a rescaled t-statistic, $t/\sqrt{T}$, for testing long-horizon regressions. Its asymptotic distribution, although non-normal, is easy to simulate. The results are quite general and applicable whenever long-horizon regressions are employed. The rescaled t-statistic is easy to use and can be computed with existing computer routines. Using arguments similar to those in Richardson and Stock (1989) and Viceira (1997), I show that its asymptotic distributions can be approximated with a relatively small sample. Phillips (1986) uses similar arguments to show that, in the context of spurious regressions, the $t/\sqrt{T}$ statistic converges to a well-defined distribution. For a good, although a bit outdated, overview of this literature, see Stock (1994).

I use the derived analytical expressions to explain the empirical and simulation results obtained by previous authors, including the excess returns/dividend yield regressions in Fama and French (1988), and the Fisher effect tests in Boudoukh and Richardson (1993), and Mishkin (1992). For example, I address the interesting question of whether long-horizon regressions have greater power to detect deviations from the null than do short-horizon regressions, or whether the significant results are a mere product of size distortion. This question, indirectly discussed in Hodrick (1992), Mishkin (1990, 1992), Goetzmann and Jorion (1993), and Campbell (2001), is posed explicitly in Campbell et al. (1997). Some Monte Carlo simulations suggest power gains (Hodrick, 1992), and others show size distortions (Goetzmann and Jorion, 1993; Nelson and Kim, 1993), but a definite, analytic answer has not yet been provided. I show that the significant results from long-horizon regressions are due to incorrect critical values. However, if appropriately rescaled, tests of long-horizon regressions have a somewhat better power at rejecting alternatives than their short-horizon analogues. Another implication of my analysis is that a significant R^2 in such regressions cannot be interpreted as an indication of a good fit.

I use the FCLT because it has been shown to provide a very good approximation of the finite-sample distributions of interest. Richardson and Stock (1989) use a similar methodology, but they consider only univariate regressions. Their results can be viewed as a special case within my framework. Viceira (1997) uses FCLT asymptotics to analyze various tests for structural breaks. In recent work, Lanne (2002) and Torous et al. (2002) also consider a particular case of long-horizon regressions, representing a special case in my framework. In this paper, I offer an exhaustive and coherent treatment of regressions in which the overlap in the observations is a nontrivial fraction of the sample size. The complete analysis allows me to compare various ways of running long-horizon regressions.

This analytical framework is broad enough to accommodate forecasting variables that, although persistent, do not have an exact unit root. Such a generalization is important in practice since many of the predictors, although highly serially correlated, must nevertheless be stationary. Also, given the sensitivity of unit-root tests to model misspecifications, it is often hard to say whether a process is truly stationary (Schwert, 1987, 1989). The cost of this generalization is the introduction of a nuisance parameter that measures deviations from the exact unit root case. I discuss three ways of dealing with this nuisance parameter.

This paper is not a condemnation of long-horizon studies. My aim is to put inference and testing using long-run regressions on a firm basis and not to rely exclusively on simulation methods. The conclusions from Monte Carlo or bootstrap studies are limited to the case study at hand, but fail to yield general insights applicable to other cases. In contrast, my analysis provides general guidelines on how to test long-horizon relations. Many researchers are aware that normal asymptotic approximations are not adequate when using overlapping variables. The reason for the poor approximations is attributed to serial correlation in the error terms. However, Monte Carlo simulations show that even after correcting for serially correlated errors, using Hansen and Hodrick (1980) or Newey-West (1987) standard errors, the small-sample distribution of the estimators and the *t*-statistics are very different from the asymptotic normal distribution (Mishkin, 1992; Goetzmann and Jorion, 1993).

The paper is structured as follows. Section 2 presents the various ways of specifying long-horizon regressions that have commonly been used in the empirical literature. Section 3 provides the main theoretical results. Testing and inference are analyzed using asymptotic methods. In Section 4, I conduct simulations to illustrate the analytical results. Section 5 applies the conclusions in Section 3 to the excess returns/dividend yield equations of Fama and French (1988) and to the long-run Fisher effect, as tested in Boudoukh and Richardson (1993) and Mishkin (1992). Section 6 concludes.

2. The model

The underlying data-generating processes are

$$Y_{t+1} = \alpha + \beta X_t + \varepsilon_{1,t+1}, \tag{1}$$

$$(1 - \phi L)b(L)X_{t+1} = \mu + \varepsilon_{2,t+1}. \tag{2}$$

The variable X_t is represented as an autoregressive process, whose highest root, ϕ, is conveniently factored out and $b(L) = b_0 + b_1 L + b_2 L^2 + \cdots + b_p L^p$ is invertible. Let $\phi = 1 + c/T$, where the parameter c measures deviations from the unit root in a decreasing (at rate T) neighborhood of 1. The unit-root case corresponds to $c = 0$. This parameterization allows us to examine highly persistent regressors, such as the dividend-price ratio, inflation, and the short interest rate, that might behave similarly to a unit-root process in a finite sample.[1] Define $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, where ε_t is a martingale difference sequence with $E(\varepsilon_t \varepsilon_t' | \varepsilon_{t-1}, \ldots) = \Sigma = [\sigma_{11}^2 \ \ \sigma_{12}; \sigma_{21} \ \ \sigma_{22}^2]$ and finite fourth moments. Finally, to alleviate notation, let $\alpha = \mu = 0$. All regressions are run with a constant term. The above assumptions can be relaxed considerably without affecting the conclusions of the paper. Note that we could have started with a general vector autoregression (VAR) as in Watson (1994). However, the system $(1-2)$ is convenient to use in the studies cited above, since we know what subset of variables have a stochastic trend. In the stock return/dividend yield example, X_t is the dividend yield, whereas Y_{t+1} is the excess return. In the inflation/interest rate

[1] For more references on local-to-unity processes, see Phillips (1987), Stock (1991), Cavanagh et al. (1995).

literature, X_t denotes the interest rate. The timing of the variables is chosen to conform exactly with the previous literature on excess returns and dividend yields (Stambaugh, 1986, 1999; Cavanagh et al., 1995), but it is not essential for the results.

Running regression (1) often yields poor results, in the sense that the OLS estimate of β is insignificant and the R^2 is extremely small. This is not surprising in the predictive regressions literature, since there is much more noise in Eq. (1) than signal coming from X_t, as discussed in Torous and Valkanov (2002). The lack of testing and explanatory power has prompted researchers to look for ways of aggregating the data in order to obtain more precise estimates. Intuitively, the aggregation of a series into a long-horizon variable is thought to strengthen the signal, while eliminating the noise. Given $(1-2)$, the long-horizon variables are

$$Z_t^k = \sum_{i=0}^{k-1} Y_{t+i},$$

$$Q_t^k = \sum_{i=0}^{k-1} X_{t+i}$$

and regressions are run in two fashions: Z_{t+1}^k on X_t, and Z_t^k on Q_t^k (or on Q_{t-k}^k). Examples of the first type are Fama and French (1988), Campbell and Shiller (1988), Lanne (2002), and Torous et al. (2002). In the predictive regression literature, Z_t^k is the kth-period continuously compounded (log) return. Examples of the second type of long-horizon regressions are Boudoukh and Richardson (1993), Mishkin (1992), and Fisher and Seater (1993), where Z_t^k is the kth-period continuously compounded (log) return and the kth period GDP growth, respectively, and Q_t^k is the kth period expected inflation and the growth (or level) of nominal money supply, respectively.

In addition to the two types of long-horizon regressions, it is often convenient to adopt different null hypotheses for β. In the predictive regressions literature, it is appropriate to assume that dividends have no predictive power for future returns, or $\beta = 0$ (Fama and French, 1988; Campbell and Shiller, 1988). In the Fisher equation literature, it is often assumed that $\beta = 1$, or that nominal interest rates move one-for-one with inflation (Mishkin, 1992; Boudoukh and Richardson, 1993).

The empirically interesting specifications of long-horizon regressions can be categorized into four cases, presented in Table 1. The regressand is always $Z_{t+1}^k = \sum_{i=0}^{k-1} Y_{t+1+i}$, but the regressor and the relation vary under the null. In the first two

Table 1
Specifications of long-horizon regressions

	Under the null	
Regressor	$\beta = 0$	$\beta \neq 0$
X_t	Case 1	Case 3
Q_t^k	Case 2	Case 4

Notes: Various ways of specifying long-horizon regressions. The regress and is always $Z_{t+1}^k = \sum_{i=0}^{k-1} Y_{t+1+i}$.

cases, one regresses Z_{t+1}^k on X_t or Q_t^k, and testing is carried out under the null of $\beta = 0$, whereas in cases 3 and 4, the null is $\beta = \beta_0 \neq 0$. The distinction is important because the stochastic behavior of Y_t and Q_t^k is different depending on β. If $\beta = 0$, Y_t is an $I(0)$ process, whereas if $\beta \neq 0$, Y_t is $I(1)$.

Let the time overlap in the summations be a fixed fraction of the sample size, or $k = [\lambda T]$. Similar parameterizations are used by Richardson and Stock (1989) in the univariate context, by Viceira (1997) in deriving tests for structural breaks, and by Valkanov (1998) in estimating the persistence of short-term interest rates. More specifically, Richardson and Stock consider case 1, where the regression is Z_t^k on Z_{t-k}^k.

Undoubtedly, the absence of an econometric foundation has greatly contributed to such a variety of ways of specifying long-horizon regressions. I prove that the statistics in long-horizon relations do not have convenient properties, namely consistent OLS estimators of the slope coefficient, t-tests with adequate power and size, and R^2 converging to one in probability under the null. More importantly, in *all* cases, the t-statistic fails to converge to a well-defined distribution. It "explodes" as the overlap increases. In finite samples, this results in increasingly significant t-tests as the overlap increases, whether or not there is a relation between the variables. Similar results can be observed in the simulations by Hodrick (1992), Goetzmann and Jorion (1993), Mishkin (1992), and Nelson and Kim (1993), but the econometric properties of the statistics were never systematically analyzed.

The reason underlying the results is simple. Overlapping a nontrivial fraction of the sample produces a persistent variable that behaves very much like an $I(1)$ process. To obtain t-statistics that have a well-defined distribution, one must divide them by the square root of the sample size.

3. Theoretical results

In this section, I present the analytical results. In addition to stating the theorems, I also provide informal discussions to clarify their implications and applications. Proofs are in Appendix A.

The assumptions and additional notation are summarized here for convenience.

Assumptions. In model $(1-2)$,

1. $Z_t^k = \sum_{i=0}^{k-1} Y_{t+i}$ and $Q_t^k = \sum_{i=0}^{k-1} X_{t+i}$.
2. The portion of overlapping is a fraction of the sample size, or $k = [\lambda T]$, where λ is fixed between 0 and 1, and [.] denotes the lesser greatest integer operator.
3. $\phi = 1 + c/T$ and $\alpha = \mu = 0$.
4. $\varepsilon_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, where ε_t is a martingale difference sequence with $\mathrm{E}(\varepsilon_t\varepsilon_t'|\varepsilon_{t-1}, \ldots) = \Sigma = [\sigma_{11}^2 \;\; \sigma_{12}; \sigma_{21} \;\; \sigma_{22}^2]$ and finite fourth moments.
5. The roots of $b(L) = b_0 + b_1L + b_2L^2 + \cdots + b_pL^p$ are less than one in absolute value, $\sum_{i=1}^{p} i|b_i| < \infty$, and p is a fixed number.

From the Functional Central Limit Theorem (FCLT), it is known that under the above assumptions, $(1/\sqrt{T}\sigma_{11} \sum_{i=0}^{[sT]} \varepsilon_{1,i}, 1/\sqrt{T}\sigma_{22} \sum_{i=0}^{[sT]} \varepsilon_{2,i}) \Rightarrow (W_1(s), W_2(s))$

jointly, where $\Rightarrow$ denotes weak convergence, and $W_1(s)$ and $W_2(s)$ are two standard Weiner processes on $\underline{D}[0,1]$, with covariance $\delta = \sigma_{12}/(\sigma_{11}\sigma_{22})$. Similarly, if $\omega^2 = \sigma_{11}^2/b^2(1)$, then $(1/\omega\sqrt{T})X_{[sT]} \Rightarrow J_c(s)$, where $J_c(s)$ is an Ornstein-Uhlenbeck process, defined by $\mathrm{d}J_c(s) = cJ_c(s) + \mathrm{d}W_2(s)$, with $J_c(0) = 0$. Sometimes, it would be more useful to write $J_c(s)$ as $J_c(s) = W_2(s) + c\int_0^s \mathrm{e}^{c(s-\tau)} W_2(\tau)\,\mathrm{d}\tau$, as in this representation the effect of the noncentrality parameter c is more evident. Note that for $c = 0, (1/\omega\sqrt{T})X_{[sT]} \Rightarrow W_2(s)$.

The assumptions above simplify the analysis but can easily be generalized. For instance, ε_t can follow a less restrictive time-series process as in Stock and Watson (1989). More generally, the results below will be valid for any strongly mixing process ε_t, thus allowing for weakly dependent and possibly heterogeneous innovations (Hansen, 1992). All regressions are run with a constant, as is usually done in practice.

3.1. Approximating the finite-sample distributions

Before deriving the main results, I prove a Lemma that will serve as a foundation for the rest of the paper. In this Lemma, I show that long-horizon variables, which are nothing but partial rescaled sums of the underlying processes X_t and Y_t, converge weakly to functionals of diffusion processes.

Lemma 1. *If Assumptions 1–5 hold, then*

1. $1/\omega T^{3/2} Q_t^k \Rightarrow \int_s^{s+\lambda} J_c(\tau)\,\mathrm{d}\tau \equiv \bar{J}_c(s;\lambda)$ *and*
2. $1/\omega T^{3/2}(Q_t^k - \bar{Q}^k) \Rightarrow \bar{J}_c(s;\lambda) - 1/(1-\lambda)\int_0^{1-\lambda} \bar{J}_c(s;\lambda)\,\mathrm{d}s \equiv \bar{J}_c^{\mu}(s;\lambda)$.
   If $\beta = 0$, *then*
3. $1/\sqrt{T}\sigma_{11} Z_t^k \Rightarrow W_1(s+\lambda) - W_1(s) = W_1(s;\lambda)$ *and*
4. $1/\sqrt{T}\sigma_{11}(Z_t^k - \bar{Z}^k) \Rightarrow W_1(s;\lambda) - 1/(1-\lambda)\int_0^{1-\lambda} W_1(s;\lambda)\,\mathrm{d}s = W_1^{\mu}(s;\lambda)$;
   and if $\beta \neq 0$, *then*
5. $1/\omega T^{3/2} Z_t^k \Rightarrow \beta\bar{J}_c(s;\lambda)$ *and*
6. $1/\omega T^{3/2}(Z_t^k - \bar{Z}^k) \Rightarrow \beta\bar{J}_c^{\mu}(s;\lambda)$.

To simplify notation, let's define the following functionals.

$$F_1(A(s),B(s)) \equiv \frac{\int_0^{1-\lambda} A(s)B(s)\,\mathrm{d}s}{\int_0^{1-\lambda}(B(s))^2\,\mathrm{d}s},$$

$$F_2(A(s),B(s)) \equiv \frac{(\int_0^{1-\lambda} A(s)B(s)\,\mathrm{d}s)}{[\int_0^{1-\lambda}(A(s))^2\,\mathrm{d}s\int_0^{1-\lambda}(B(s))^2\,\mathrm{d}s - (\int_0^{1-\lambda} A(s)B(s)\,\mathrm{d}s)^2]^{1/2}},$$

$$F_3(A(s),B(s)) \equiv \frac{(\int_0^{1-\lambda}(A(s)B(s))\,\mathrm{d}s)^2}{\int_0^{1-\lambda}(A(s))^2\,\mathrm{d}s\int_0^{1-\lambda}(B(s))^2\,\mathrm{d}s}.$$

All distributions can be represented as one of these functionals, using the diffusion processes in Lemma 1 as their arguments.

We start by presenting the results for case 1.

Theorem 2. *When* $\beta = 0$ *and Assumptions 1–5 hold, if we regress* Z_{t+1}^k *on a constant and* X_t*, the slope coefficient and the associated statistics will have the following properties*:

- $\hat{\beta} \Rightarrow \sigma_{11}/\omega F_1(W_1^\mu(s;\lambda), J_c^\mu(s))$.
- $t_{\hat{\beta}}/T^{1/2} \Rightarrow F_2(W_1^\mu(s;\lambda), J_c^\mu(s))$.
- $R^2 \Rightarrow F_3(W_1^\mu(s;\lambda), J_c^\mu(s))$.

Some results are worth emphasizing. First, $\hat{\beta}$ is not a consistent estimator of β, and its distribution depends on σ_{11} and ω, which can be estimated consistently from Eqs. (1) and (2). However, the limiting distribution also depends on the nuisance parameter c, which cannot be estimated consistently using only the time-series properties of the data. Section 3.2 discusses three ways of dealing with c. Second, the t-statistic does not converge to a well-defined distribution, but diverges at rate $T^{1/2}$. We cannot rely on asymptotic values to construct correctly sized confidence intervals. In other words, a bigger sample size or a bigger overlap (since $k = [\lambda T]$) will tend to produce higher t-statistics. Therefore, we can account for the results in Fama and French (1988) and Campbell and Shiller (1988), where the t-statistics are increasing with the horizon of the regression. One way around this problem is to carry out simulations on a case-by-case basis. A more general method of testing the slope coefficient of long-horizon regressions is to use the $t/\sqrt{T}$ statistic. As we will see below, this statistic converges weakly in all four cases. Moreover, its distribution is simple to simulate and depends only on the parameters c and δ, which can either be estimated consistently as discussed above, or marginalized by using sub-bound intervals. Lastly, the coefficient of determination R^2 does not converge in probability under the null, thus explaining why aggregating the data tends to produce high R^2.

Theorem 3. *When* $\beta = 0$ *and Assumptions 1–5 hold, if we regress* Z_{t+1}^k *on a constant and* Q_t^k*, the slope coefficient and the associated statistics will have the following properties*:

- $T(\hat{\beta} - 0) \Rightarrow \sigma_{11}/\omega F_1(W_1^\mu(s;\lambda), \bar{J}_c^\mu(s;\lambda))$,
- $t/T^{1/2} \Rightarrow F_2(W_1^\mu(s;\lambda), \bar{J}_c^\mu(s;\lambda))$,
- $R^2 \Rightarrow F_3(W_1^\mu(s;\lambda), \bar{J}_c^\mu(s;\lambda))$.

This is case 2. Here $\hat{\beta}$ is a consistent estimator of β. In fact, it converges to the true value of zero at rate T, which is higher than the usual $\sqrt{T}$ convergence that obtains with stationary regressors. However, the t-statistic does not converge to a well-defined distribution. Similar to case 1, the R^2 does not converge in probability. In fact, most of the discussion from case 1 is also applicable here. It is important to notice that the $t/\sqrt{T}$ statistic converges weakly. Testing can be carried out by simulating its limiting distribution and calculating its asymptotic critical values.

Theorem 4. *When $\beta \neq 0$ and Assumptions 1–5 hold, if we regress Z_{t+1}^k on a constant and X_t, the slope coefficient and the associated statistics will have the following properties*:

- $\hat{\beta}/T \Rightarrow \beta F_1(J_c^\mu(s), \bar{J}_c^\mu(s;\lambda))$,
- $t_{\hat{\beta}}/T^{1/2} \Rightarrow F_2(J_c^\mu(s), \bar{J}_c^\mu(s;\lambda))$,
- $R^2 \Rightarrow F_3(J_c^\mu(s), \bar{J}_c^\mu(s;\lambda))$.

Theorem 4 states that, in case 3, $\hat{\beta}$ is not a consistent estimator for β and does not have a well-defined distribution. As the overlap increases, one would tend to observe slope coefficients that are increasing in magnitude, as is the case in Fama and French (1988), Campbell and Shiller (1988), and Boudoukh and Richardson (1993). Moreover, the limiting distribution of $\hat{\beta}/T$ depends on the unknown parameter β itself. Similar to cases 1 and 2, the t-statistic must be normalized by the square root of the sample size to converge to a well-defined distribution. The limiting distribution depends only on one nuisance parameter, c (but not on δ). The coefficient of determination does not converge in probability, but has a well-defined distribution. However, it cannot be used to judge the fit of the regression in the usual fashion.

Lastly, we present the results for case 4.

Theorem 5. *When $\beta \neq 0$ and Assumptions 1–5 hold, if we regress Z_{t+1}^k on a constant and Q_t^k, the slope coefficient and the associated statistics will have the following properties*:

- $T(\hat{\beta} - \beta) \Rightarrow \sigma_{11}/\omega F_1(W_1^\mu(s), \bar{J}_c^\mu(s;\lambda))$,
- $t_{\hat{\beta}}/T^{1/2} \Rightarrow F_2(W_1^\mu(s), \bar{J}_c^\mu(s;\lambda))$,
- $R^2 \rightarrow^p 1$.

In many respects, this is econometrically the most desirable way of estimating a long-run regression. First, the estimator is super-consistent. Second, after appropriate normalization, the t-statistic converges to a distribution that can easily be simulated, provided we have consistent estimates of c and δ. Third, unlike in the previous three cases, the R^2 converges in probability to one.

An interesting pattern emerges from the results above. Regressing a long-run variable on a short-run variable, as in cases 1 and 3, yields inconsistent estimators of the slope coefficients, whether or not there is a true relation between Y_{t+1} and X_t. However, projecting a long-horizon variable on another long-horizon variable, as in cases 2 and 4, produces super-consistent estimators of the true parameter, whether it be $\beta = 0$ or $\beta = \beta_0 \neq 0$. Moreover, the limiting distribution of $\hat{\beta}$ and $t/\sqrt{T}$ is very similar for cases 2 and 4.

Table 2 provides a quick summary of the above results. The stochastic order notation $V_T = O_p(1)$ intuitively indicates that the variable V_T has a well-defined distribution as $T \rightarrow \infty$. In all four cases, inference based on the t-statistic cannot be made using asymptotic critical values. However, the normalized t-statistic, $t/\sqrt{T}$, has a well-defined distribution that can be easily simulated. As Monte Carlo

Table 2
Specifications of long-horizon regressions: results

Regressor	Under the null			
	$\beta = 0$		$\beta \neq 0$	
X_t	Case 1	$\hat{\beta} = O_p(1)$ $\frac{t}{T^{1/2}} = O_p(1)$ $R^2 = O_p(1)$	Case 3	$\frac{\hat{\beta}}{T} = O_p(1)$ $\frac{t}{T^{1/2}} = O_p(1)$ $R^2 = O_p(1)$
$\sum_{i=0}^{k-1} X_{t-k+i}$	Case 2	$T(\hat{\beta} - 0) = O_p(1)$ $\frac{t}{T^{1/2}} = O_p(1)$ $R^2 = O_p(1)$	Case 4	$T(\hat{\beta} - \beta) = O_p(1)$ $\frac{t}{T^{1/2}} = O_p(1)$ $R^2 \to^p 1$

Notes: Summary of the results in Theorems 2–5.

simulations would show, the convergence is quite fast. Hence, asymptotic critical values can be calculated using a relatively small sample. To obtain accurate estimates of the slope parameter, one should run long-horizon variables on long-horizon variables, as in cases 2 and 4, where the estimators are super-consistent. Lastly, the R^2 should not be trusted as a measure of regression fit in the usual sense. This statistic converges to one only in Theorem 5. In the other cases, it can lead to false conclusions, as demonstrated in the next section.

In essence, there are two distinct ways of conducting long-horizon regressions: Z_{t+1}^k on X_t and Z_{t+1}^k on Q_t. Often, the specification of the regression is motivated by theory, as in the case of returns and dividend yields. Other times, the regression is motived by a "looser" desire to investigate the relation between two variables at longer horizons. Ultimately, the goal is to find out whether the maintained null hypothesis is true, and one cannot help but wonder which setup, Z_{t+1}^k on X_t or Z_{t+1}^k on Q_t, would yield a more powerful test for this hypothesis. We know that the $t/\sqrt{T}$ statistics in cases 1 and 2 converge to different limiting distributions. It is clear that the power of the tests must be different. However, an analytical comparison of their power functions is not possible. As noted earlier, the limiting distributions of $t/\sqrt{T}$ depend on the nuisance parameter c. Although consistent estimators of c can often be designed by using additional modeling assumptions, it is sometimes time-consuming or impossible to do so. In such cases, one would be compelled to use strictly statistical procedures for marginalizing c, perhaps by constructing conservative sup-bound intervals. The cost of this simplification would be a loss in power. Comparing the power functions of $t/\sqrt{T}$ in cases 1 and 2, with a consistent estimate of c and using sup-bound intervals, is addressed by simulations in the next section.

3.2. The nuisance parameter c

To simulate the above distributions, we need to either assume that c is known or estimate it. However, by construction, c cannot be estimated consistently from a unique realization of the time-series. In this section, I discuss three ways of dealing with this problem. The first two methods, from Stock (1991) and Cavanagh et al. (1995), are well known and only briefly described.

Stock (1991) proposes an asymptotically median-unbiased estimator of c, c^{MUE}, and centered confidence intervals by inverting the distribution of the augmented Dickey–Filler statistic. Such a method is easy to implement with a computer. The main drawback of Stock's (1991) procedure is that the inversion of quantiles of a distribution usually yields confidence intervals that are too wide for practical purposes. Such intervals often comprise a wide range of alternatives, and in particular $c = 0$, suggesting that the forecasting variable might be nonstationary (Stock, 1991). This criticism will also be valid in our applications.

Cavanagh et al. (1995) propose to circumvent the estimation of the nuisance parameter by taking the extreme values of the test statistics of interest over a wide range of c. Such a test, called a sup-bound test, produces conservative confidence intervals.[2] This method is quite general, since it relies only on statistical arguments. However, in addition to conservative size, the sup-bound test will also have low power of rejecting false alternatives, since it does not use any information about the magnitude of c (Section 4.2).

Valkanov (1998) suggests using the long-run restriction implied by an economic model in order to consistently estimate c. For example, in the case of returns and dividend yields, one might use the dynamic Gordon growth model in Campbell and Shiller (1988) to estimate c, as I do in Section 5.1. Appendix B provides a more detailed discussion of this new method. Unlike the above procedures, such an estimator, denoted by $c^{\text{CONSISTENT}}$, will not only be consistent but it will converge at rate T, thus providing very precise point estimates of the nuisance parameter. Conditional on the validity of the model, using $c^{\text{CONSISTENT}}$ in order to compute the $t/\sqrt{T}$ statistic will yield correctly sized tests with a higher power function than the sup-bound tests. Since the three methods of dealing with the nuisance parameter are not equivalent, they are all applied (in Section 5) in the testing for the relations of interest.

4. Simulations

The theorems in the previous section provide an asymptotic approximation of the distributions of $\hat{\beta}, t/\sqrt{T}$, and R^2. It is well known that rescaled partial sums converge quickly to their limiting distributions (Stock, 1994). In other words, the asymptotic distributions provide a very accurate approximation of the small-sample distributions even for samples of (say) 100 observations. Here, I conduct Monte Carlo simulations to illustrate some of the points made in the previous section. First,

[2] More formally, let $d_{c,\alpha}$ denote the $100\alpha\%$ quantile of the distribution $F_2(.,.)$ for a given value of δ. Then, define $(\underline{d}_\alpha, \bar{d}_\alpha) = (\inf_c d_{c,\alpha}, \sup_c d_{c,\alpha})$. Then a conservative, say 95%, sub-bound test is one that rejects $t/\sqrt{T} \notin (\underline{d}_{0.025}, \bar{d}_{0.975})$. Cavanagh et al. (1995) also use Bonferroni and Scheffe-type intervals, in addition to the sub-bound intervals, to marginalize c. None of the three methods necessarily dominates the other two. The sub-bound test has the virtue of simplicity and, as shown in Cavanagh et al. (1995) it performs no worse than the other two methods.

I demonstrate that the $t/\sqrt{T}$ statistic converges asymptotically for all of the above cases, unlike the non-rescaled t-statistic. Second, I plot the densities of the t-statistics derived in Theorems 2–5 and demonstrate the accuracy of the approximations. Third, I investigate how these densities depend on the nuisance parameters c and δ. Last, I investigate the power properties of the proposed $t/\sqrt{T}$ test for cases 1 and 2, with and without a consistent estimator of c.

4.1. Validity of the asymptotic approximations

The experiment is conducted as follows. For every case in Table 1, I simulate series of 100 and 750 observations 5,000 times using Eqs. (1) and (2). The first sample length corresponds to a typical annual data set, whereas the second corresponds to a series with monthly frequencies. The residuals $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are normally distributed, with a correlation of $\delta \in \{0, -0.9\}$.[3] Let $c \in \{0, -0.5, -1, -5, -10\}$. For each simulation, I compute $\hat{\beta}$, the t-statistic under the appropriate null hypothesis, and R^2. For simplicity, I set $\sigma_{11}^2 = \omega^2 = 1$ without losing generality, since the asymptotic distributions of $t/\sqrt{T}$ (and R^2) are invariant to those parameters.

The simulated densities of the t-statistics are plotted in Fig. 1 using values of the nuisance parameters $c = -1$ and $\delta = 0$. The first column of graphs displays the densities of the non-normalized t-statistics, whereas the second column displays the densities of the appropriately rescaled statistics, as suggested by Theorems 2–5. The variances of the t-statistics increase with the sample size, in all four cases. There are no correct asymptotic critical values for the non-rescaled t-statistics. For $T = 750$, the variance is considerable. Thus, it is not surprising that Mishkin's (1992) simulations lead him to conclude that "the t-statistics need to be greater than 14 to indicate a statistically significant β coefficient..." (Mishkin, 1992, p. 203). Goetzmann and Jorion (1993) use the bootstrap method to reach a similar conclusion: "OLS t-statistics over 18 and R^2 over 38% for all multiple-year horizons are not unusual." However, the rescaled t-statistic, $t/\sqrt{T}$, converges to a well-defined distribution. Notice that the partial sums converge very fast to their limiting distributions, as the densities simulated with 100 and 750 observations are almost identical. The properties of $\hat{\beta}$ and R^2 from Theorems 2–5 can be illustrated with similar figures (omitted for conciseness.)

The results so far plot the distributions of the t-statistic for specific values of the pair (c, δ). To understand the dependence of the distribution on the nuisance parameters, Table 3 provides the first two simulated moments of this statistic in cases 1–4, for $c \in \{0, -0.5, -1, -5, -10\}$ and $\delta \in \{0, -0.9\}$. The results are similar to the $c = -1$, $\delta = 0$ case. The column "Ratio" displays the variance of the statistic for $T = 750$ divided by the variance for $T = 100$. If the ratio of the variances is equal to one, then the estimated second moment of the distribution has not changed with the increased sample size. This ratio can be used to verify the rates of convergence of our statistics. For instance, in Table 3, case 1, the ratio of the variances of the t-statistic,

[3] The results hold true for various specifications of the residuals. Simulations using various serially correlated and conditionally heteroskedastic errors are available upon request.

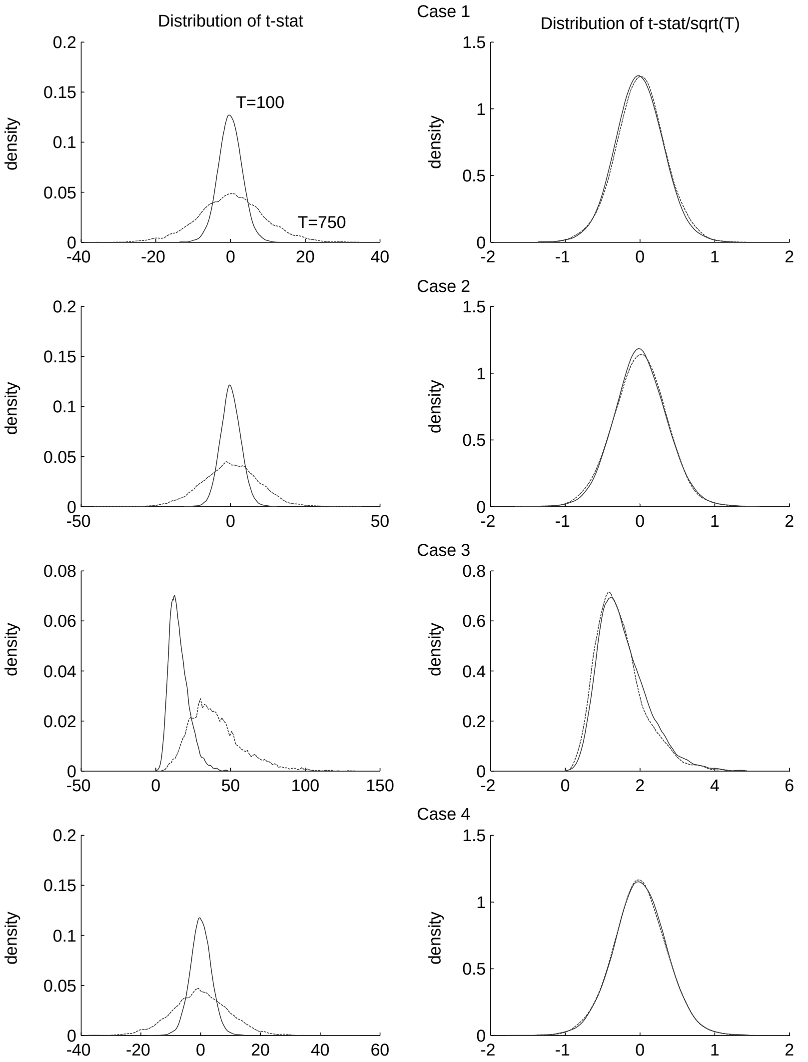

Fig. 1. Simulation results of t-statistics. *Notes*: Figure 1 plots the simulated densities of the non-rescaled and rescaled t-statistics in all four cases. The graphs in the first column are the distributions of the t-statistic under the null $\beta = 0$, $c = -1$, $\delta = 0$, and $T = 100$ and $T = 750$. The distributions of the same statistic, appropriately rescaled to converge asymptotically, are presented in the second column. For an exact description of the simulations, see the text and Table 3. The distributions for $T = 100$ are almost indistinguishable from those with $T = 750$, suggesting a fast convergence of the distribution of $t/\sqrt{T}$. Therefore, the above methods are particularly appropriate for analyzing small-sample distributions of the statistics of interest. The variance of the t-statistic increases with the sample size, but $t/\sqrt{T}$ has a well-defined asymptotic distribution, as expected from Theorems 2–5.

for $c = 0$, $\delta = 0$, is 7.54, indicating that an increase in the sample size by a factor of 7.5 has also resulted in an increase of the variance of the t-statistic by a factor of 7.54. This result is in precise agreement with Theorem 2.

Table 3
Simulations of the distribution of t-statistics

	$\hat{E}(t-stat)$				$\hat{Var}(t-stat)$					
	$\delta=0.00$		$\delta=-0.90$		$\delta=0.00$			$\delta=-0.90$		
c	$T=100$	$T=750$	$T=100$	$T=750$	$T=100$	$T=750$	Ratio	$T=100$	$T=750$	Ratio
Case 1										
0.0	−0.078	−0.043	3.480	12.289	10.616	80.033	7.54	7.483	73.142	9.77
−0.5	0.032	−0.047	3.118	11.405	10.513	79.716	7.58	7.099	65.193	9.18
−1.0	−0.092	0.254	2.791	10.801	10.064	78.946	7.84	6.570	64.789	9.86
−5.0	0.011	0.012	1.198	7.011	8.834	67.546	7.65	5.758	55.286	9.60
−10.0	−0.005	−0.082	0.269	4.846	7.831	57.135	7.30	5.355	50.243	9.38
Case 2										
0.0	−0.051	0.210	−0.429	−0.211	11.723	91.043	7.77	5.968	45.952	7.70
−0.5	0.054	0.058	−1.006	−1.409	11.546	92.545	8.02	5.283	37.597	7.12
−1.0	0.073	0.069	−1.477	−2.869	11.941	92.148	7.72	4.682	37.396	7.99
−5.0	0.025	−0.247	−4.372	−9.916	11.930	87.706	7.35	4.550	35.534	7.81
−10.0	−0.062	−0.184	−6.624	−15.510	10.650	82.782	7.77	5.772	40.865	7.08
Case 3										
0.0	18.884	48.544	22.571	49.687	93.003	646.840	6.96	117.792	651.176	5.53
−0.5	16.983	43.396	20.361	44.337	62.999	433.632	6.88	77.409	446.140	5.76
−1.0	15.546	40.125	18.795	40.633	44.162	316.620	7.17	57.017	329.765	5.78
−5.0	11.043	27.879	13.578	28.547	16.868	117.041	6.94	19.836	113.668	5.73
−10.0	8.448	21.310	10.590	22.123	9.100	61.564	6.77	9.616	61.902	6.44
Case 4										
0.0	0.023	0.069	−0.522	−0.153	12.249	92.081	7.52	5.857	44.590	7.61
−0.5	0.011	−0.060	−0.992	−1.494	11.997	96.491	8.04	5.165	39.425	7.63
−1.0	0.035	−0.060	−1.535	−2.957	12.288	92.233	7.51	4.796	37.674	7.86
−5.0	0.009	−0.042	−4.286	−9.946	12.175	86.888	7.14	4.683	34.287	7.32
−10.0	−0.049	−0.164	−6.634	−15.556	11.335	85.948	7.58	6.033	40.198	6.66

Notes: The table displays the first two moments of the non-normalized t-statistic under the null for various values of c, δ, and T. The data were generated as follows: $Y_t = \alpha + \beta X_{t-1} + \varepsilon_{1,t},\ X_t = \phi X_{t-1} + \varepsilon_{2,t}$, where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are standard normal variates with correlation $\delta = (0, -0.9)$, $\theta = 1 + c/T$, for $c = (0, -0.5, -1, -5, -10)$, and $T = (100, 750)$. The case $c = 0$ represents the exact unit-root. In cases 1 and 2, we let $\beta = 0$, whereas in cases 3 and 4, $\beta = 1$. The data are simulated 5,000 times. For each simulation, I construct $Z^k_{t+1} = \sum_{i=0}^k Y_{t+i}$ and, in cases 2 and 4, $Q^k_t = \sum_{i=0}^k X_{t+i}$ to be the long-horizon regressand and regressor, respectively. I let $k = [\lambda T]$ and $\lambda = 0.1$. The tabulated results are those from running Z^k_{t+1} on X_t (for cases 1 and 3) and Z^k_{t+1} on Q^k_t (for cases 2 and 4). In each table, the column "Ratio" presents the variance of the corresponding statistic for $T = 750$ divided by the same variance for $T = 100$. This ratio allows us to check the rates of convergence of the t-statistics. For example, in the first panel we can see that an increase in the sample from 100 to 750 observations results in an increase of the variance of the t-statistic by approximately a factor of 7.5, exactly as dictated by Theorem 2. The same holds true for the other three panels. Therefore a non-normalized t-statistic would not be suitable for testing.

It is also important to know how sensitive the tails of the distributions are with respect to the nuisance parameters. Table 4 displays the quantiles 1, 2.5, 5, 95, 97.5, and 99 of the $t/\sqrt{T}$ statistic for various values of the parameters c and δ in all four

Table 4
Tails of $t/\sqrt{T}$ statistic at various percentiles

	$\delta = 0.00$						$\delta = -0.90$					
c	1.0	2.5	5.0	95.0	97.5	99.0	1.0	2.5	5.0	95.0	97.5	99.0
Case 1												
0.0	−0.771	−0.659	−0.538	0.543	0.661	0.773	−0.267	−0.151	−0.051	0.975	1.088	1.237
−0.5	−0.818	−0.681	−0.532	0.544	0.642	0.774	−0.297	−0.169	−0.070	0.907	1.002	1.110
−1.0	−0.772	−0.639	−0.526	0.548	0.670	0.787	−0.282	−0.175	−0.087	0.870	0.982	1.130
−5.0	−0.727	−0.583	−0.496	0.499	0.593	0.718	−0.384	−0.268	−0.176	0.706	0.805	0.922
−10.0	−0.649	−0.533	−0.450	0.442	0.541	0.643	−0.405	−0.316	−0.244	0.608	0.686	0.801
Case 2												
0.0	−0.797	−0.672	−0.563	0.595	0.727	0.853	−0.619	−0.508	−0.419	0.390	0.481	0.612
−0.5	−0.841	−0.688	−0.572	0.579	0.717	0.883	−0.653	−0.510	−0.408	0.308	0.391	0.494
−1.0	−0.817	−0.698	−0.577	0.564	0.682	0.855	−0.695	−0.585	−0.485	0.233	0.318	0.407
−5.0	−0.862	−0.698	−0.580	0.533	0.667	0.810	−0.984	−0.854	−0.743	−0.036	0.016	0.085
−10.0	−0.829	−0.660	−0.545	0.543	0.651	0.771	−1.182	−1.071	−0.974	−0.209	−0.146	−0.065
Case 3												
0.0	0.406	0.510	0.636	3.586	4.045	4.811	0.492	0.594	0.689	3.610	4.092	4.693
−0.5	0.396	0.519	0.625	3.046	3.473	3.983	0.439	0.555	0.659	3.091	3.570	4.070
−1.0	0.377	0.495	0.610	2.705	3.014	3.524	0.436	0.539	0.623	2.761	3.087	3.475
−5.0	0.288	0.380	0.449	1.732	1.896	2.152	0.329	0.424	0.485	1.753	1.915	2.117
−10.0	0.225	0.292	0.350	1.292	1.426	1.539	0.228	0.308	0.380	1.312	1.432	1.599
Case 4												
0.0	−0.834	−0.700	−0.564	0.574	0.692	0.835	−0.589	−0.502	−0.410	0.396	0.495	0.609
−0.5	−0.855	−0.715	−0.583	0.582	0.712	0.893	−0.653	−0.523	−0.433	0.320	0.395	0.492
−1.0	−0.833	−0.707	−0.575	0.568	0.696	0.855	−0.683	−0.572	−0.490	0.247	0.321	0.433
−5.0	−0.847	−0.669	−0.553	0.547	0.675	0.808	−0.960	−0.826	−0.737	−0.040	0.022	0.113
−10.0	−0.846	−0.689	−0.567	0.552	0.662	0.793	−1.177	−1.064	−0.976	−0.215	−0.151	−0.079

Notes: Table 4 displays the tail probabilities of the distribution of $t/\sqrt{T}$ for cases 1–4. For an exact description of the simulations, see Section 4 or Table 3 above. Since the distributions depend on the nuisance parameters c and δ, I present percentiles 1, 2.5, 5, 95, 97.5, and 99 of the distributions for various values of those parameters. The distributions were simulated with samples of length $T = 750$. In all four cases, the distributions are very similar for small values of c ($c \leqslant -5$). For larger values of c, the distributions are similar only when the correlation between the residuals is small. This implies that, for persistent processes, the size distortions induced by the unknown parameter c will not be large.

cases. As expected, the distribution of $t/\sqrt{T}$ changes noticeably in cases 1, 2, and 4 for $\delta \in \{0, -0.9\}$. In case 3, the distribution is invariant to δ, as shown in Theorem 4. Given that we can obtain a consistent estimate of δ, the dependence of the distributions on this parameter does not represent a concern. The tails of the distributions are less sensitive to different values of c. In fact, for values of the local-to-unity parameter close to zero, say $c \leqslant -5$, the percentiles are very similar in all four cases, especially for $\delta = 0$. From Table 4, we draw the conclusion that the distribution of the $t/\sqrt{T}$ statistic will be insensitive to estimation errors in c and δ, as long as those parameters are close to zero. For values δ close to -1, we need accurate estimates of both nuisance parameters.

4.2. Power comparisons

The last set of simulations is designed to compare the power functions of various tests of the null $\beta = 0$. The data under the null is simulated as described in the previous section. I investigate the power of the $t/\sqrt{T}$ statistic by running regressions of Z_{t+1}^k on X_t (case 1) and Z_{t+1}^k on Q_t^k (case 2), where $k = [\lambda T]$ and $\lambda = 0.1$. Since the null $\beta = 0$ can also be tested using the short-horizon regression (1), I consider the usual t-statistic of such a regression, where the standard error of the coefficient is computed using the Newey–West standard errors. To compute the power function of the three tests, I simulate the model under a set of alternatives $\beta \in (0, 0.1)$.

There are three interesting questions to be answered. The first is which of the three tests is more powerful, given that we have a consistent estimate of c. The second is what test is the most powerful, given that we have to resort to sup-bound intervals. The third is how big is the loss in power when we use the conservative sup-bound intervals. All simulations are conducted under the unfavorable scenario of $\delta = -0.9$ since, as seen in Table 4, a high correlation will insure that the exact and the conservative confidence intervals have a different coverage.

The results for the case $T = 100$ and $c = -1$ are plotted in Fig. 2. The first plot displays the power function for local alternatives, i.e., alternatives that are close to the null, whereas the second plot shows the entire power function. Given a consistent estimate of c (solid lines), the power of the $t/\sqrt{T}$ statistic is similar under both long-horizon specifications. For alternatives away from the null, the $t/\sqrt{T}$ test in the long-horizon regressions offers significant power gains compared to the t-test in the short-horizon regression. A similar conclusion holds true if a consistent estimate of c is not available (dashed lines). In such a case, the $\sup_c t/\sqrt{T}$ tests clearly dominate the short-horizon $\sup_c t$ test. Moreover, the power function of the sup-bound intervals in case 1 dominates the power function of case 2. Different horizons k (or λ) yield almost identical results. There is a considerable loss in power from not having a consistent estimate of c if δ is high in absolute value. To summarize the results, the $t/\sqrt{T}$ statistic is able to reject false hypotheses with high probability, especially when a precise estimate of c is available. An accurate estimate of c is also indispensable when the correlation between the residuals is high.

5. Empirical results: the excess returns/dividend yield and Fisher effect regressions

5.1. Long-horizon predictability of excess returns using dividend yields

The predictability of excess returns, labeled as one the "new facts in finance" by Cochrane (1999), is so widely accepted in the profession that it has generated a new wave of models (e.g., Barberis, 2000; Brennan et al., 1997; Campbell and Viceira, 1999; Liu, 1999) that try to analyze the implications of return predictability on

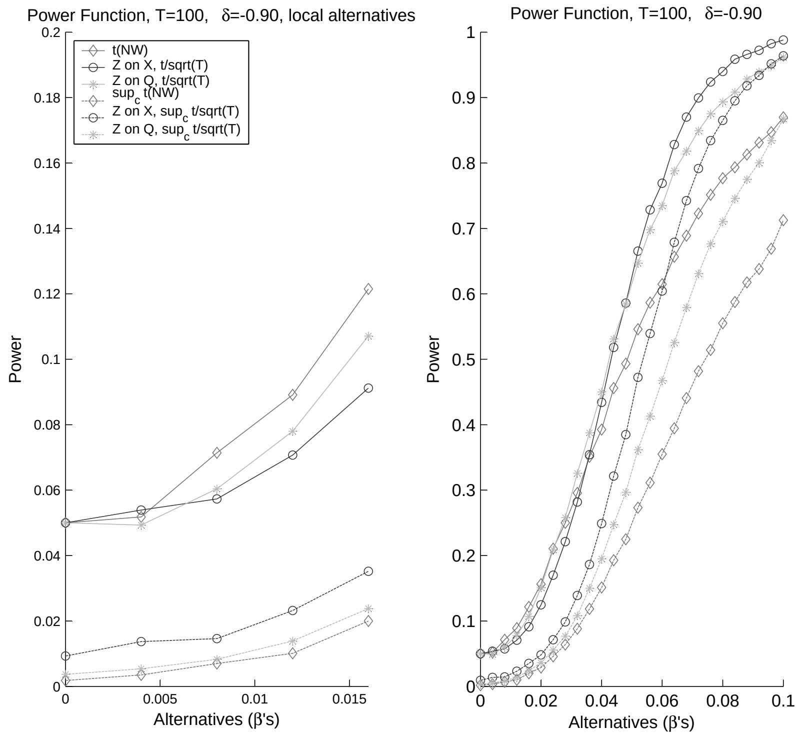

Fig. 2. Power comparison of various tests. *Notes*: The data under the null $\beta = 0$ are simulated as described in Section 4.1 and in Table 3. The data under the alternatives are simulated for $\beta \in (0, 0.1)$. I consider the power functions of the statistic $t/\sqrt{T}$ from the regression of Z_{t+1}^k on X_t (case 1) and Z_{t+1}^k on Q_t^k (case 1), where $k = [\lambda T]$ and $\lambda = 0.1$. The results for the case $T = 100$ and $c = -1$ are plotted above. The power function of the Newey–West t-test in a short-horizon regression of Y_{t+1} on X_t is also plotted, as a reference. The first plot displays the power function for local alternatives, i.e., alternatives that are close to the null, whereas the second plot shows the entire power function. Solid lines signify power functions computed using a consistent estimate of c. Dotted lines signify that the tests were conducted by forming conservative sup-bound intervals, as in Cavanagh et al. (1995). Given a consistent estimate of c, the power of the $t/\sqrt{T}$ statistic is similar under both long-horizon specifications. For alternatives away from the null, the $t/\sqrt{T}$ test in the long-horizon regressions offers significant power gains compared to the t-test in the short-horizon regression. If a consistent estimate of c is not available, then the $\sup_c t/\sqrt{T}$ test from the regression of Z_{t+1}^k on X_t dominates the $\sup_c t/\sqrt{T}$ test from the regression of Z_{t+1}^k on Q_t^k, which in turn dominates the $\sup_c t$ test. The power loss from not having an estimate of c is noticeable, but not dramatic.

portfolio decisions. Fama and French (1988) and Campbell and Shiller (1988) first argued that, unlike short-horizon returns, long-horizon returns can be predicted using dividend yields or dividend-price ratios. The predictability is measured in-sample from the t-statistic and the R^2. Both of these statistics increase with the horizon (the overlap). However, Goetzmann and Jorion (1993) and Nelson and Kim (1993) conduct simulations and bootstrap studies on the properties of the estimates from the excess returns/dividend yield long-horizon regressions and conclude that the evidence for predictability is not nearly as overwhelming at long horizons as

Table 5
Summary statistics

Panel A Series	$\bar{X}$	$\hat{\sigma}$	$\hat{\phi}$	P-value(ADF)
VW log(D/P)	−3.23	0.33	0.99	0.37
EW log(D/P)	−3.23	0.33	0.99	0.39
VW(1)	6.46	19.20	0.10	0.00
EW(1)	8.76	25.32	0.18	0.00
VW(12)	6.31	20.98	0.93	0.00
EW(12)	8.64	28.44	0.94	0.00
VW(24)	6.14	20.75	0.97	0.96
EW(24)	8.44	27.99	0.97	0.94
VW(36)	5.98	19.84	0.98	0.85
EW(36)	8.32	26.60	0.98	0.81
VW(48)	5.98	19.13	0.98	0.77
EW(48)	8.40	25.75	0.98	0.78
VW(60)	6.11	18.17	0.98	0.82
EW(60)	8.59	23.94	0.98	0.84
VW(72)	6.34	16.36	0.98	0.87
EW(72)	8.84	20.93	0.98	0.86
CPI Infl	3.11	1.96	0.91	0.25
3M Tbill	3.71	0.91	0.97	0.74

Panel B *Correlation of residuals* (δ)	
VW $Corr(\varepsilon_1, \varepsilon_2)$	−0.94
EW $Corr(\varepsilon_1, \varepsilon_2)$	−0.85
Infl. $Corr(\varepsilon_1, \varepsilon_2)$	0.04
Int.Rate $Corr(\varepsilon_1, \varepsilon_2)$	0.06

Notes: Panel A presents the summary statistics of the variables of interest. The first two columns display the means and standard deviations. The third column is the highest autoregressive root of the series, whereas the fourth column presents the p-value of the Augmented Dickey Fuller test under the null of a unit root in the series. The first two series are the log dividend price ratios of the CRSP value-weighted and equal-weighted portfolios. The next 14 series are the continuously compounded value-weighted and equal-weighted CRSP excess returns, where the period of the return is shown in parentheses. For example, VW(24) is the 24-month continuously compounded excess value-weighted return computed using overlapping one-period returns. All returns are expressed in annualized percentage points. The last two series are the CPI inflation and the nominal interest rate of a three-month T-bill. Panel B presents the correlation of the residuals, as discussed in the text.

previous studies had suggested. Yet, Hodrick (1992) uses VAR simulation-based studies to conclude that the distortions in the distributions of the t-statistics were not enough to overturn the results of Fama and French (1988). It is difficult to reconcile or compare the conclusions from these studies, because they are all based on simulations or bootstrap methods, and fail to yield general results. Their conclusions are ultimately a function of how the artificial data are being generated. More important, there is no way of knowing what features of the data would influence the statistics of interest and in what fashion.

Section 3 provided a framework which will now be applied to explain and re-test the results from Fama and French (1988). I take the monthly value- and equal-weighted returns and the corresponding dividend yields from the Center for Research in Security Prices (CRSP) dataset from 1927 to 1999. In accord with most recent studies, I work with excess log returns (log returns minus the log yield of the three-month Treasury bill), represented in percentage points (i.e. multiplied by 100). The long-horizon continuously compounded excess returns are formed by taking a rolling sum of the one-period log excess returns over a given time interval. The annual dividend yield ratio is obtained by dividing the sum of the dividends over the last year by the current price. To keep the regressions consistent with the log-linearized framework of Campbell and Shiller (1988), I work with log dividend yields (see Campbell et al., 1997). Summary statistics of the series are shown in Table 5. The lag structure of the dividend yield series is chosen with sequential t-tests, as suggested by Ng and Perron (1995). The selected lag structure has five lags, yielding highest autoregressive roots of 0.987 and 0.989 for the log dividend price ratio in the value-weighted and equal-weighted portfolios, respectively (Table 5). It is interesting to point out that an augmented Dickey–Fuller (ADF) test cannot reject the null of a unit root in the log dividend price ratio, but neither can it reject alternatives close to the null (i.e. local-to-unity alternatives). I have chosen the ADF test since, as pointed out by Schwert (1987, 1989), it is less sensitive to model misspecifications than other unit-root tests. Finally, I obtain an estimate of $\hat{\delta} = -0.94$ for the value-weighted equation and $\hat{\delta} = -0.85$ for the equal-weighted regression, using the residuals from regressions $(1-2)$. Keeping in mind that in the case of highly serially correlated residuals the distribution of the $t/\sqrt{T}$ statistic is sensitive to estimation errors in c (Table 4), we need to find a precise estimate of that parameter.

Table 6 displays the median unbiased estimate, $\hat{c}^{\mathrm{MUE}}$, and 95%-centered confidence intervals by inverting the distribution of the ADF test. A consistent estimate of c, $\hat{c}^{\mathrm{CONSISTENT}}$, is also computed following the methods discussed in Section 3.2 and Appendix B. Although $\hat{c}^{\mathrm{MUE}}$ and $\hat{c}^{\mathrm{CONSISTENT}}$ are similar in three of the subsamples, there are some significant differences. Notice that three out of the four confidence intervals around $\hat{c}^{\mathrm{MUE}}$ include the explosive, unit-root case, which is untenable from a theoretical perspective. It is hard to think of a theoretical reasoning that might justify a unit root in the dividend yield and a unit-root component in returns; see Campbell et al. (1997). In the most recent period, $\hat{c}^{\mathrm{MUE}}$ is positive, suggesting that log dividend yield is an explosive process. In contrast, $\hat{c}^{\mathrm{CONSISTENT}}$ is always negative and its 95% confidence intervals are always below (although barely) the unit-root threshold. Moreover, the confidence intervals around $\hat{c}^{\mathrm{CONSISTENT}}$ are

Table 6
Estimates of c and ϕ in the dividend yield process

	Samples			
	1927–1999	1927–1957	1957–1987	1957–1999
$\hat{c}^{\text{MUE}}$	−0.754	−36.243	−0.970	0.079
	(−20.798,0.093)	(−50.000,−0.908)	(−35.196,0.009)	(−1.375,0.525)
$\hat{\phi}$	0.9991	0.8993	0.9973	1.0002
	(0.9759,1.0001)	(0.8333,0.9975)	(0.9022,1.0000)	(0.9973,1.0010)
$\hat{c}^{\text{CONSISTENT}}$	−1.401	−0.728	−0.662	−0.206
	(−1.606,−1.196)	(−0.895,−0.560)	(−0.798,−0.526)	(−0.378,−0.034)
$\hat{\phi}$	0.9983	0.9977	0.9978	0.9995
	(0.9978,0.9987)	(0.9973,0.9980)	(0.9975,0.9981)	(0.9992,0.9999)

Notes: The first and second rows provide the point estimate and the 95% confidence interval, respectively, of the parameters c and ϕ, where $X_t = \phi X_{t-1} + \varepsilon_{2,t}$, $\phi = 1 + c/T$, and X_t is the dividend yield. The median unbiased estimate of c, $\hat{c}^{\text{MUE}}$, is obtained by inverting the 50th percentile of the augmented Dickey-Fuller test (ADF) with 12 lags (Stock, 1991), chosen with sequential pre-testing. The 95% confidence interval, displayed below the estimate, is obtained in a similar fashion, by inverting percentiles 2.5nd and 97.5th of the ADF test. The corresponding value of ϕ and its confidence intervals are also displayed for convenience. A consistent estimate of c, $\hat{c}^{\text{CONSISTENT}}$, obtained by using the theoretical relation between log returns and log dividend yields, its 95% confidence interval, and the corresponding values of ϕ, are displayed in the third and fourth rows. As discussed in the text, this estimator has theoretical advantages over the median unbiased estimator. It also performs better in practice. Notice that three out of the four confidence intervals around $\hat{c}^{\text{MUE}}$ include the unit root case, which is unappealing from a theoretical and practical perspective, as discussed in the text. Moreover, in the last subsample, the MUE of c is positive, suggesting that the log dividend yield is a non-stationary process. In contrast, the consistent estimates of c are all negative and their confidence intervals are all below, although barely, the unit root threshold.

extremely tight, suggesting the efficiency gains from using the information provided by the dynamic Gordon growth model. In this application, I will use $\hat{c}^{\text{CONSISTENT}}$, given its superiority.

Table 7 presents the results from the long-horizon regressions for the value- and equal-weighted excess returns, for different sample periods and different horizons. The results are very familiar from the work of Fama and French (1988) and Campbell et al. (1997). The difference in magnitude is because I work with returns expressed as percentage points. The high slope estimates, the (Newey–West) t-statistics that increase with the horizon, and the R^2's that increase with the horizon should not come as a surprise, as argued in Section 2. Here, I focus on the renormalized $t/\sqrt{T}$ test, since as proved in Theorem 2 above, it converges to a well-defined distribution. The estimated values of the $t/\sqrt{T}$ statistic and the corresponding p-value under the null $\beta = 0$ (case 1, above) are shown in the last two columns of each table. The distribution of $t/\sqrt{T}$ under the null is simulated with the estimated values of $(\hat{c}^{\text{CONSISTENT}}, \hat{\delta})$ and $\lambda = k/T$, where k is the horizon of the regression. The results stand in stark contrast to those obtained with conventional inference methods. Using the entire sample period, 1927–1999, we cannot reject the null of no predictability at usual levels of significance for any of the horizons.

Table 7
Excess returns/dividend yield predictive regressions

	Excess VW returns					Excess EW returns				
Horizon	$\hat{\beta}$	$t(NW)$	R^2	$t/\sqrt{T}$	P-value	$\hat{\beta}$	$t(NW)$	R^2	$t/\sqrt{T}$	P-value
Period: 1927–1999										
1	0.55	0.69	0.00	0.03	0.000	1.57	1.57	0.00	0.07	0.00
12	9.97	1.48	0.02	0.16	0.300	25.50	2.92	0.09	0.31	0.02
24	21.42	2.17	0.06	0.25	0.300	48.98	4.02	0.16	0.44	0.03
36	29.90	2.43	0.08	0.30	0.350	63.00	4.20	0.20	0.50	0.07
48	38.47	2.54	0.11	0.35	0.320	75.57	3.88	0.23	0.55	0.09
60	47.28	2.90	0.15	0.41	0.290	85.63	4.39	0.28	0.62	0.10
Period: 1927–1945										
1	0.41	0.14	0.00	0.01	0.010	2.55	0.70	0.00	0.06	0.00
12	22.48	1.20	0.04	0.19	0.600	56.49	2.23	0.13	0.38	0.21
24	68.42	3.04	0.16	0.43	0.420	133.25	4.38	0.34	0.71	0.14
36	105.50	2.48	0.27	0.61	0.360	175.05	3.36	0.42	0.85	0.15
48	137.96	3.43	0.39	0.79	0.290	209.13	5.01	0.49	0.97	0.17
60	145.40	4.19	0.44	0.88	0.290	212.80	7.33	0.53	1.05	0.20
Period: 1946–1980										
1	2.43	3.53	0.02	0.15	0.000	2.52	2.74	0.02	0.12	0.00
12	30.84	4.51	0.26	0.59	0.000	34.53	3.16	0.18	0.47	0.01
24	56.41	4.66	0.45	0.91	0.010	58.52	2.82	0.26	0.58	0.08
36	72.84	5.60	0.57	1.16	0.010	68.30	2.67	0.26	0.60	0.18
48	84.38	6.28	0.61	1.26	0.010	71.63	2.51	0.24	0.56	0.31
60	103.49	6.48	0.63	1.30	0.030	81.94	2.65	0.24	0.56	0.39
Period: 1946–1999										
1	0.82	1.42	0.00	0.06	0.000	1.40	2.06	0.01	0.08	0.00
12	10.71	1.93	0.05	0.23	0.250	20.07	3.27	0.09	0.32	0.07
24	18.79	1.76	0.09	0.31	0.340	33.50	3.38	0.15	0.42	0.16
36	25.57	1.88	0.12	0.37	0.370	42.56	3.78	0.18	0.46	0.24
48	32.24	2.02	0.15	0.43	0.380	49.66	4.09	0.20	0.49	0.31
60	44.12	2.37	0.20	0.50	0.360	59.66	4.44	0.22	0.52	0.33

Notes: The regressor is the continuously compounded excess value- or equal-weighted return for different periods, shown in the first column. The regressand is the lagged log dividend/price ratio. Results using contemporaneous log dividend/price ratios are almost identical and hence omitted. All regressions are run with a constant. The second column displays the OLS estimate of β. The Newey–West t-statistic and the coefficient of determination, R^2, are shown in columns 3 and 4. Column 5 displays the $t/\sqrt{T}$ statistic, and the last column displays the p-value, using this normalized t-statistic under the null $\beta = 0$. As discussed in the test, the $t/\sqrt{T}$ statistic has a well-defined asymptotic distribution. Its distribution is simulated with the consistent estimates of c, or $c = (-1.40, -0.73, -0.66, -0.21)$ (see Table 6) for the corresponding sample periods, using $\lambda = k/T$, where k is the horizon of the regression. The correlation between the residuals is set to equal the sample estimates, or -0.94 in the value-weighted regressions, and -0.85 for the equal-weighted regressions. Approximate critical values of the $t/\sqrt{T}$ statistic can be obtained from Table 4, case 1, for $c = -1$ and $\delta = -0.9$. The results are very similar if we use median unbiased estimates of c, or $c = (-0.75, -36.20, -0.97, 0.08)$ (see Table 6). The 95% sup-bound (conservative) confidence interval for $t/\sqrt{T}$ is: $[-0.316, 1.088]$ for $\lambda = 0.1$ and the extreme values are taken over the range $c = (0, -10)$. The coverage of the interval does not change significantly for other horizons.

Looking at the various subsamples, the lack of predictability seems to come mainly from the 1927–1945 and 1981–1999 periods. If there is any evidence of return predictability, it is strongest during the 1946–1980 sample for the value-weighted returns, a result that has also been pointed out by Viceira (1997). During that period, the evidence of predictability is statistically significant at the 5% level. Moreover, the statistics from the equal-weighted returns indicate that the forecasting relation between the log dividend price ratio and returns (of mainly small companies) was strongest during the 1927–1945 period and has been declining ever since. Viceira conducts a more complete analysis of the changing excess returns/dividend yield relation using Functional Central Limit Theory. The main conclusion to be retained from Table 5 is that when appropriate testing procedures are employed, the evidence of return predictability is not as strong as previously maintained.

5.2. Long-horizon Fisher effect

I focus on two of the most recent tests of the Fisher effect, namely the long-horizon approaches of Mishkin (1992) and Boudoukh and Richardson (1993). Neither paper can reject the null that long-horizon returns and long-horizon inflation move one for one (this is known as the full Fisher effect), whereas previous studies have often had trouble finding any positive correlation between short-horizon returns and inflation. In light of the theoretical discussion above, could it be that the results in those two studies are due to the fact that long-horizon returns and long-horizon inflation are constructed by summing over short-horizon variables?

5.2.1. Nominal stock returns and inflation

Boudoukh and Richardson focus on the regression $\sum_{i=1}^{5} R_{t+i} = \alpha_5 + \beta_5 \sum_{i=1}^{5} \pi_{t+i} + \varepsilon_t$, where R_t is the stock return, π_t is the inflation rate, and the null is $\beta_5 = 1$. This long-horizon regression corresponds to case 4. The tests in the original study are conducted with standard normal asymptotic critical values. To remedy measurement problems that might arise from using ex post instead of ex ante inflation, the authors also estimate the above regression using several instruments, one of which is lagged long-horizon inflation, $\sum_{i=1}^{5} \pi_{t-5+i}$. I revisit the OLS and the instrumental variable (IV) calculations using the estimates from Boudoukh and Richardson, Tables 1 and 2. The normalized t-statistic is computed by dividing the reported t-statistic by the square root of the sample size and its p-value is obtained from simulations. The results are presented in Panel A of Table 8 for the entire sample and for two subsamples.

I obtain an estimate of $\hat{\delta} = 0.04$ (see Table 5) using the data from Siegel (1992) and Schwert (1990) and simulate the limiting distribution from Theorem 5. The estimates of δ for the subsamples are very close to the one from the entire sample. Since there is no reasons to suspect variations in this parameter, I use the more precise estimate from the 1802–1990 period. In this application, given that no appropriate model is available to consistently estimate c, I use $\hat{c}^{\mathrm{MUE}}$. As discussed in the previous section, when δ is small (in absolute value), the limiting distribution of the $t/\sqrt{T}$ statistic will vary little with c. I experimented with other values of the local-to-unity parameter,

Table 8
Fisher's equation

Horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	P-value
Panel A: stock returns				
Period: 1802–1990				
5 years	0.524	−2.736	−0.200	0.27
IV: Lagged 5 years	1.820	0.752	0.055	0.56
Period: 1870–1990				
5 years	0.462	−2.093	−0.191	0.28
IV: Lagged 5 years	1.434	0.559	0.041	0.55
Period: 1914–1990				
5 years	0.432	−11.360	−1.303	0.01
IV: Lagged 5 years	2.120	0.941	0.069	0.58
Panel B: interest rates				
Period: 1953–1990				
3 months	0.663	−5.507	−0.261	0.21
Period: 1953–1979				
3 months	1.188	2.642	0.150	0.66
Period: 1979–1990				
3 months	0.235	−3.027	−0.505	0.07
Period: 1982–1990				
3 months	0.125	−5.573	−0.569	0.05

Notes: Panel A of Table 8 reports the OLS estimate of regressing the five-month nominal stock return on the five-month inflation rate. The values in the first row of each table are taken from Boudoukh and Richardson (1993), Table 1. The values in the second row labeled "IV: Lagged 5 years" are taken from Boudoukh and Richardson (1993), Table 2, case ii. The third column is the appropriately normalized $t/\sqrt{T}$ statistic. The last column reports the percentile of the $t/\sqrt{T}$ statistic, under the null of a full Fisher effect (Mishkin, 1992; Gali, 1988), or $\beta = 1$. This long-horizon regression corresponds to case 4. The distribution of the normalized t-statistic is simulated with $\hat{\delta} = 0.04$ and $\hat{c}^{\mathrm{MUE}} = (-9, -12, -7)$, where c is estimated using Stock's (1991) median unbiased estimator for each period. The test cannot reject the null in two out of the three periods for the OLS case, contrary to the conclusions reached from the non-normalized t-test. For the IV case, the null cannot be rejected in any period. The 95% sub-bound confidence interval is $[-0.715, 0.712]$ for the OLS case and $[-0.725, 0.718]$ for the IV case. The first two columns in Panel B of Table 8 report the OLS estimate of regressing the three-month inflation rate on the three-month nominal interest rate and the t-statistic under the null $\beta = 1$. The values are taken from Mishkin (1992), Table 1. The last two columns present the normalized $t/\sqrt{T}$ statistic and its percentile under the null of a full Fisher Effect. The nuisance parameters δ and c are set to equal their corresponding estimates, or $\hat{\delta} = 0.06$ and $\hat{c}^{\mathrm{CONSISTENT}} = -1$, where the last estimate is taken from Valkanov (1998). This long-horizon regression corresponds to case 4. Unlike the t-statistic, the $t/\sqrt{T}$ statistic cannot reject the null. However, the evidence for a full Fisher effect is less convincing for the last two subperiods. A similar conclusion was reached by Mishkin (Tables 3–4) using Monte Carlo simulations. The 95% sup-bound (conservative) confidence interval is $[-0.712, 0.711]$.

and the results were almost identical. In fact, the sup-bound intervals were very similar to those obtained with $\hat{c}^{\text{MUE}}$. The percentiles of the $t/\sqrt{T}$ statistic, computed under the null $\beta = 1$, are reported in column 4. Note that the statistics in the IV case will have a slightly different asymptotic distribution because of the lags. The results from Theorem 4 for the IV case are: $T(\hat{\beta} - \beta) \Rightarrow \sigma_{11}/\omega F_1(W_1^{\mu}(s;\lambda), \bar{W}_2^{\mu}(s - \lambda; s))t_{\hat{\beta}}/T^{1/2} \Rightarrow F_2(W_1^{\mu}(s;\lambda), \bar{W}_2^{\mu}(s-\lambda;s))$, and $R^2 \to^p 1$.

Rejection would occur if the statistic fell in the tails of the distribution. Values lower than 0.025 or higher than 0.975 indicate rejection at the 5% level. For the OLS case, the null of a full Fisher effect cannot be rejected at the 5% level, except for the 1914–1990 subperiod, where it cannot be rejected at the 1% level. This conclusion contradicts the one from the usual t-statistics, where the rejection seems to be categorical in all periods. In the IV case, the null cannot be rejected for any period. I also performed tests under the null $\beta = 0$, corresponding to case 1. The null is rejected at all levels of significance, thus providing an indirect indication of good testing power. To conclude, a re-examination of the evidence in Boudoukh and Richardson (1993) supports a Fisher effect in the OLS and IV case.

5.2.2. *Inflation and nominal interest rates*

Mishkin (1992) focuses on a similar regression using monthly macroeconomic data described in Mishkin (1990). The regressor is the three-period interest rate, whereas the regressand is the three-period inflation.[4] Despite the relatively small overlap in the creation of the new series, Monte Carlo simulations conducted by the author (Tables 3–4) convinced me that my asymptotic approximations can appropriately be used in this case. As mentioned above, Mishkin (1992) concludes that the correct 5% critical value for the t-test is 14 for the entire sample, and 20 for the 1953–1979 subsample. Recalling Fig. 1, those values are in accord with what one would expect in case 4 when using the non-rescaled t-statistic. In fact, Mishkin remarks that the 'potential for a spurious regression result between the level of interest rates and future inflation is thus very high' (Mishkin, 1992, p. 203). However, he fails to notice the possibility of rejecting the null of a full Fisher effect, $\beta = 1$, when it is true, as a direct result of creating long-horizon variables. Indeed, $\beta = 1$ is rejected using conventional critical values (Mishkin, 1992, Table 1).

We test the null $\beta = 1$ by rescaling the t-statistic, reported in Mishkin (1992), and simulating its limiting distribution according to Theorem 5. The results are reported in Panel B of Table 8, for the period 1953–1990 and different subperiods. The last column shows the percentile of the computed $t/\sqrt{T}$ statistic under the null, and for $\delta = 0.06$ (see Table 5). I take the estimate of $\hat{c}^{\text{CONSISTENT}}$, obtained in Valkanov (1998), for the same short interest rate process.[5] The estimate is in the range $(-3, -1)$ depending on the sample period. However, as pointed out in Section 4, the limiting

[4] Let $p_t = \log(P_t)$, where P_t is the price level at time t. The three-period inflation is computed as $\pi_t^3 = p_t - p_{t-3} = p_t - p_{t-1} + p_{t-1} - p_{t-2} + p_{t-2} - p_{t-3} = \pi_t^1 + \pi_{t-1}^1 + \pi_{t-2}^1$. More generally, $(1 - L^k)Y_{t+k} = (1 - L)(1 + L + \cdots + L^{k-1})Y_{t+k} = \sum_{i=1}^{k} \Delta Y_{t+i}$.

[5] Valkanov (1998) uses the expectations hypothesis of the term structure as a way of imposing a long-horizon restriction.

distribution of $t/\sqrt{T}$ will be "almost" invariant to different values of c due to the small correlation δ.

If inference is conducted using the t-statistic (column 2) and standard normal critical values, the null would be rejected in all periods at usual significance levels. Rejection occurs because the t-statistic does not converge asymptotically to a well-defined distribution (Theorem 5). If we compute the $t/\sqrt{T}$ statistic, the null cannot be rejected at the 5% level (two-sided test) for any of the subperiods. The evidence for a full Fisher effect is weaker, but still significant, in the post-1979 period. Mishkin (1992) reaches the same conclusion by conducting a series of Monte Carlo simulations (his Table 4), even though he does not provide an econometric explanation of the results.

The $t/\sqrt{T}$ tests of a long-horizon Fisher effect in the stock (Boudoukh and Richardson, 1993) and bond (Mishkin, 1992) markets remarkably lead us to the same conclusion. There is strong evidence of a full Fisher effect for the periods before 1979. In the post-1979 era, the evidence is still present, but not as convincing.

6. Conclusion

I analyze four ways of conducting long-horizon regressions that have frequently been used in empirical finance and macroeconomics. The least squares estimator of the slope coefficient, its t-statistic, and the R^2 have non-standard asymptotic properties. I reach several conclusions. First, the coefficient is not always consistently estimated. For reliable estimates, one must specify regressions by aggregating both the regressor and the regressand (cases 2 and 4), i.e., running one long-horizon variable against another. Second, the standard t-statistic does not converge asymptotically to a well-defined distribution in any of the cases. The practical implication is that an increase in the horizon of the regression will result in higher t-values. Therefore, testing cannot be conducted using the customary standard normal critical values, since it would most likely lead to rejecting the null very often, even when it is true. In order to conduct asymptotically valid tests, I propose the $t/\sqrt{T}$ statistic, which has the virtue of being easily computed. Its limiting distribution, although non-normal, is fast to converge, easy to simulate, and depends on only one nuisance parameter that can be estimated consistently. Third, the R^2 in long-horizon regressions does not converge in probability under the null in three of the cases. Therefore, it cannot be interpreted as a measure of the goodness of fit in the regression. Fourth, I provide a precise and consistent estimate of the nuisance parameter c, which plays a crucial role in all distributions, particularly when the correlation between the residuals is high.

The above results are applicable whenever long-horizon regressions are used. The tendency of long-run methods to produce "significant" results, no matter what the null hypothesis, should neither come as a surprise, nor be taken as conclusive evidence. In light of the present arguments, the tests in long-horizon studies must be re-evaluated. The last section of this paper employs the proposed $t/\sqrt{T}$ statistic to re-examine the predictability of returns in Fama and French (1988) using dividend yields and the dividend-price ratio. I find little predictability for the periods before World War II. In the post-war period, the dividend-price ratio and the dividend yield seem to have some predictive power, but the evidence is not as striking as previously maintained. In another application, I re-visit the conclusions from the long-run Fisher effect literature. While there is strong evidence for a full Fisher effect during the periods before 1979, the post-1979 results are not entirely convincing.

Appendix A

Proof of Lemma 1. Recall that $1/\omega T^{1/2}X_t \Rightarrow J_c(s)$, where from now on $t=[sT]$ and $\omega^2=\sigma_{22}^2/b(1)$. Also, recall that $Q_t^k=\sum_{i=0}^{k-1} X_{t+i}$. Letting $k=[\lambda T]$, we can write $1/\omega T^{3/2}Q_t^k = 1/T\ \sum_{i=0}^{k-1} X_{t+i}/\omega T^{1/2} \Rightarrow \int_s^{s+\lambda} \bar{J}_c(\tau)\,\mathrm{d}\tau \equiv \bar{J}_c(s;\lambda)$, using the continuous mapping theorem (CMT) to prove part 1. Similarly, for part 2, $1/\omega T^{3/2}\bar{Q}^k = 1/(T-k)\sum_{t=1}^{T-k+1} 1/\omega T^{3/2}Q_t^k \Rightarrow 1/(1-\lambda)\int_0^{1-\lambda} \bar{J}_c(s;\lambda)\,\mathrm{d}s$, using the CMT. Then, we can write, $1/\omega T^{3/2}(Q_t^k-\bar{Q}^k) \Rightarrow \bar{J}_c(s;\lambda) - 1/(1-\lambda)\int_0^{1-\lambda}\bar{J}_c(s;\lambda)$ $\mathrm{d}s \equiv \bar{J}_c^{\mu}(s;\lambda)$.

If $\beta=0$, then $Y_t=\varepsilon_{1,t}$. Recall that $Z_t^k=\sum_{i=0}^{k-1} Y_{t+i+1}$ and $1/\sigma_{11}T^{1/2}Z_t^k = 1/\sigma_{11}T^{1/2}\ \sum_{i=0}^{k-1}\ \varepsilon_{1,t+i+1} = 1/\sigma_{11}T^{1/2}\{\sum_{i=1}^{t+k}\varepsilon_{1,i}-\sum_{i=1}^{t}\varepsilon_{1,i}\} \Rightarrow W_1(s+\lambda)-W_1(s) \equiv W_1(s;\lambda)$, finishing part 3. Similarly, $1/\sigma_{11}T^{1/2}\bar{Z}^k = 1/(T-k)\sum_{t=1}^{T-k} 1/\sigma_{11}T^{1/2}Z_t^k \Rightarrow 1/(1-\lambda)\int_0^{1-\lambda} W_1(s;\lambda)\,\mathrm{d}s$, and $1/\sigma_{11}T^{1/2}(Z_t^k-\bar{Z}^k) \Rightarrow W_1(s;\lambda) - 1/(1-\lambda)\int_0^{1-\lambda} W_1$ $(s;\lambda)\,\mathrm{d}s \equiv W_1^{\mu}(s;\lambda)$, thus completing the proof of part 4.

When $\beta\neq 0$, $Y_t=\beta X_{t-1}+\varepsilon_{1,t}$. Therefore, $1/\omega T^{3/2}Z_t^k = 1/\omega T^{3/2}\ \sum_{i=0}^{k-1}\ Y_{t+i+1} = \beta 1/\omega T^{3/2}\ \sum_{i=0}^{k-1}\ X_{t+i} + 1/\omega T^{3/2}\ \sum_{i=0}^{k-1}\ \varepsilon_{1,t+i} \Rightarrow \beta\bar{J}_c(s;\lambda)$ using part 1 and $1/\omega T^{3/2}\ \sum_{i=0}^{k-1}\ \varepsilon_{1,t+i} = \mathrm{o_p}(1)$. Part 6 is proven in exactly the same fashion. □

Proof of Theorem 2. By definition,

$$\hat{\beta} = \frac{\sum_{t=1}^{T-k}(Z_{t+1}^k-\bar{Z}^k)(X_t-\bar{X})}{\sum_{t=1}^{T-k}(X_t-\bar{X})^2}$$

$$= \frac{1/T\ \sum_{t=1}^{T-k}((Z_{t+1}^k-\bar{Z}^k)/T^{1/2})((X_t-\bar{X})/T^{1/2})}{1/T\ \sum_{t=1}^{T-k}((X_t-\bar{X})/T^{1/2})^2}$$

$$\Rightarrow \frac{\sigma_{11}}{\omega}\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)J_c^{\mu}(s)\,\mathrm{d}s}{\int_0^{1-\lambda}(J_2^{\mu}(s))^2\,\mathrm{d}s}$$

$$= \frac{\sigma_{11}}{\omega}F_1(W_1^{\mu}(s;\lambda), J_c^{\mu}(s))$$

using Lemma 1, part 4 and the CMT. Under the null, the t-statistic is

$$t = \frac{(\hat{\beta}-0)(\sum_{t=1}^{T-k}(X_t-\bar{X})^2)^{1/2}}{s}.$$

In our case,

$$\begin{aligned}\frac{t}{T^{1/2}} &= \frac{\hat{\beta}(1/T\ \sum_{t=1}^{T-k}((X_t-\bar{X})/T^{1/2})^2)^{1/2}}{(s^2/T)^{1/2}}\\ &\Rightarrow \frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)J_c^{\mu}(s)\,\mathrm{d}s}{[\int_0^{1-\lambda}(W_1^{\mu}(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(J_c^{\mu}(s))^2\,\mathrm{d}s-(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)J_c^{\mu}(s)\,\mathrm{d}s)^2]^{1/2}}\\ &= F_2(W_1^{\mu}(s;\lambda),J_c^{\mu}(s)).\end{aligned}$$

The coefficient of determination is defined as

$$\begin{aligned}R^2 &= \hat{\beta}\frac{\sum_{t=1}^{T-k}(X_t-\bar{X})^2}{\sum_{t=1}^{T-k}(Z_{t+1}^k-\bar{Z}^k)^2}\\ &\Rightarrow \frac{(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)J_c^{\mu}(s)\,\mathrm{d}s)^2}{\int_0^{1-\lambda}(W_1^{\mu}(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(J_c^{\mu}(s))^2\,\mathrm{d}s}\\ &= F_3(W_1^{\mu}(s;\lambda),W_2^{\mu}(s)).\qquad\square\end{aligned}$$

Proof of Theorem 3. The OLS estimator is

$$\hat{\beta}=\frac{\sum_{t=1}^{T-k}(Z_{t+1}^k-\bar{Z}^k)(Q_t^k-\bar{Q}^k)}{\sum_{t=1}^{T-k}(Q_t^k-\bar{Q}^k)^2}.$$

Using the results in Lemma 1 and the CMT

$$\begin{aligned}T(\hat{\beta}-0) &= \frac{1/T\ \sum_{t=1}^{T-k}((Z_{t+1}^k-\bar{Z}^k)/T^{1/2})((Q_t^k-\bar{Q}^k)/T^2)}{1/T\ \sum_{t=1}^{T-k}((Q_t^k-\bar{Q}^k)/T^{3/2})^2}\\ &\Rightarrow \frac{\sigma_{11}}{\omega}\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s}{\int_0^{1-\lambda}(\bar{J}_c^{\mu}(s;\lambda))^2\,\mathrm{d}s}\\ &= \frac{\sigma_{11}}{\omega}F_1(W_1^{\mu}(s;\lambda),\bar{J}_c^{\mu}(s;\lambda))\end{aligned}$$

as required.

The t-statistic under the null is

$$t=\frac{(\hat{\beta}-0)(\sum_{t=1}^{T-k}(Q_t^k-\bar{Q}^k)^2)^{1/2}}{s}\quad\text{and}$$

$$\begin{aligned}\frac{t}{T^{1/2}} &= \frac{T(\hat{\beta}-0)(1/T\ \sum_{t=1}^{T-k}((Q_t^k-\bar{Q}^k)/T^{3/2})^2)^{1/2}}{(s^2/T)^{1/2}}\\ &\Rightarrow \frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s}{[\int_0^{1-\lambda}(W_1^{\mu}(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(\bar{J}_c^{\mu}(s;\lambda))^2\,\mathrm{d}s-(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s)^2]^{1/2}}\\ &= F_2(W_1^{\mu}(s;\lambda),\bar{J}_c^{\mu}(s;\lambda)).\end{aligned}$$

In a similar fashion,

$$R^2 = \hat{\beta}\frac{\sum_{t=1t}^{T-k}(Q_t^k - \bar{Q}^k)^2}{\sum_{t=1}^{T-k}(Z_{t+1}^k - \bar{Z}^k)^2} \Rightarrow \frac{(\int_0^{1-\lambda} W_1^\mu(s;\lambda)\bar{J}_c^\mu(s;\lambda)\,\mathrm{d}s)^2}{\int_0^{1-\lambda}(W_1^\mu(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(\bar{J}_c^\mu(s;\lambda))^2\,\mathrm{d}s}$$
$$= F_3(W_1^\mu(s;\lambda), \bar{J}_c^\mu(s;\lambda)),$$

completing the proof. □

Proof of Theorem 4. The proofs follow exactly the same pattern. The OLS estimator is

$$\hat{\beta} = \frac{\sum_{t=1}^{T-k}(Z_{t+1}^k - \bar{Z}^k)(X_t - \bar{X})}{\sum_{t=1}^{T-k}(X_t - \bar{X})^2}.$$

However, since $\beta \neq 0$, we have $(Z_{t+1}^k - \bar{Z}^k)/T^{3/2} \Rightarrow \beta\bar{J}_c^\mu(s;\lambda)$ from part 6 of Lemma 1. Then,

$$\frac{\beta}{T} = \frac{1/T\ \sum_{t=1}^{T-k}((Z_{t+1}^k - \bar{Z}^k)/T^{3/2})((X_t - \bar{X})/T^{1/2})}{1/T\ \sum_{t=1}^{T-k}((X_t - \bar{X})/T^{1/2})^2} \Rightarrow \beta\frac{\int_0^{1-\lambda}\bar{J}_c^\mu(s;\lambda)J_c^\mu(s)\,\mathrm{d}s}{\int_0^{1-\lambda}(W_2^\mu(s;\lambda))^2\,\mathrm{d}s}.$$

Under the null of $\beta = \beta_0$, the t-statistic is,

$$t = \frac{(\hat{\beta} - \beta_0)(\sum_{t=1}^{T-k}(X_t - \bar{X})^2)^{1/2}}{s}.$$

We have to normalize it by $T^{1/2}$ to get

$$\frac{t}{\sqrt{T}} = \frac{((\hat{\beta} - \beta)/T)(1/T\ \sum_{t=1}^{T-k}((X_t - \bar{X})/T^{1/2})^2)^{1/2}}{(s^2/T^3)^{1/2}}$$
$$\Rightarrow \frac{\int_0^{1-\lambda} W_2^\mu(s;\lambda)\bar{J}_c^\mu(s;\lambda)\,\mathrm{d}s}{[\int_0^{1-\lambda}(W_2^\mu(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(\bar{J}_c^\mu(s;\lambda))^2\,\mathrm{d}s - (\int_0^{1-\lambda} W_2^\mu(s;\lambda)\bar{J}_c^\mu(s;\lambda)\,\mathrm{d}s)^2]^{1/2}}$$
$$= F_2(J_c^\mu(s), \bar{J}_c^\mu(s;\lambda)).$$

Finally,

$$R^2 = (\hat{\beta})^2\frac{\sum_{t=1t}^{T-k}(X_t - \bar{X})^2}{\sum_{t=1}^{T-k}(Z_{t+1}^k - \bar{Z}^k)^2}$$
$$= \left(\frac{\hat{\beta}}{T}\right)^2\frac{1/T\ \sum_{t=1t}^{T-k}((X_t - \bar{X})/T^{1/2})^2}{1/T\ \sum_{t=1}^{T-k}((Z_{t+1}^k - \bar{Z}^k)/T^{3/2})^2}$$
$$\Rightarrow \frac{(\int_0^{1-\lambda} W_2^\mu(s;\lambda)\bar{J}_c^\mu(s;\lambda)\,\mathrm{d}s)^2}{\int_0^{1-\lambda}(W_2^\mu(s;\lambda))^2\,\mathrm{d}s\int_0^{1-\lambda}(\bar{J}_c^\mu(s;\lambda))^2\,\mathrm{d}s}$$
$$= F_3(J_c^\mu(s), \bar{J}_c^\mu(s;\lambda)).\qquad \square$$

Proof of Theorem 5. This case deserves more attention. First, notice that $Z_{t+1}^k - \bar{Z}^k = \beta(Q_t^k - \bar{Q}^k) + R_t - \bar{R}$, where $R_t = \sum_{i=0}^{k-1}\varepsilon_{t+i+1}$ and $\bar{R} = 1/T\ \sum_{t=0}^{T-k} R_t$. Also,

$(R_t - \bar{R})/T^{1/2} \Rightarrow W_1^{\mu}(s;\lambda)$. The OLS estimator is

$$\hat{\beta} = \frac{\sum_{t=1}^{T-k}(Z_{t+1}^k - \bar{Z}^k)(Q_t^k - \bar{Q}^k)}{\sum_{t=1}^{T-k}(Q_t^k - \bar{Q}^k)^2} = \beta + \frac{\sum_{t=1}^{T-k}(R_t^k - \bar{R})(Q_t^k - \bar{Q}^k)}{\sum_{t=1}^{T-k}(Q_t^k - \bar{Q}^k)^2}.$$

Therefore,

$$\begin{aligned} T(\hat{\beta} - \beta) &= \frac{1/T\ \sum_{t=1}^{T-k}((R_t^k - \bar{R})/T^{1/2})((Q_t^k - \bar{Q}^k)/T^{3/2})}{1/T\ \sum_{t=1}^{T-k}((Q_t^k - \bar{Q}^k)/T^{3/2})^2} \\ &\Rightarrow \frac{\sigma_{11}}{\omega} \frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s}{\int_0^{1-\lambda}(\bar{J}_c^{\mu}(s;\lambda))^2\,\mathrm{d}s} \\ &= \frac{\sigma_{11}}{\omega} F_1(W_1^{\mu}(s;\lambda), \bar{J}_c^{\mu}(s;\lambda)). \end{aligned}$$

Using similar steps, we can show that

$$\begin{aligned} \frac{t}{\sqrt{T}} &\Rightarrow \frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s}{[\int_0^{1-\lambda}(W_1^{\mu}(s;\lambda))^2\,\mathrm{d}s \int_0^{1-\lambda}(\bar{J}_c^{\mu}(s;\lambda))^2\,\mathrm{d}s - (\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\bar{J}_c^{\mu}(s;\lambda)\,\mathrm{d}s)^2]^{1/2}} \\ &= F_2(W_1^{\mu}(s;\lambda), \bar{J}_c^{\mu}(s;\lambda)). \end{aligned}$$

Lastly,

$$R^2 = \frac{\hat{\beta}^2\ \sum_{t=1}^{T-k}(Q_t^k - \bar{Q})^2}{\beta^2\ \sum_{t=1}^{T-k}(Q_t^k - \bar{Q})^2 + 2\beta\ \sum_{t=1}^{T-k}(Q_t^k - \bar{Q})(R_t^k - \bar{R}) + \sum_{t=1}^{T-k}(R_t^k - \bar{R})^2}.$$

Dividing the numerator and the denominator by $\beta^2\ \sum_{t=1}^{T-k}(Q_t^k - \bar{Q})^2$, we obtain

$$R^2 = \frac{\hat{\beta}^2/\beta^2}{1 + o_p(1) + o_p(1)} \rightarrow^p 1. \qquad \square$$

Appendix B

Valkanov (1998) proposes to use information provided by an economic model in order to consistently estimate c. The procedure in Valkanov (1998), although yielding the most precise estimates, is conditioned upon having a model that establishes a long-horizon restriction. Since Campbell and Shiller (1988) provide such a relation in the excess returns/dividend yield case, I use their framework to estimate c.

The model used to estimate c is the log-linearized dynamic Gordon model used in Campbell and Shiller (1988), who show that the log stock return, r_{t+1}, can be written as $r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t$, with p and d denoting the log stock price and log dividend, respectively, and k and ρ are parameters of the linearization.[6] We may rearrange this expression as $r_{t+1} \approx k - \rho x_{t+1} + x_t + \Delta d_{t+1}$, where $x \equiv d - p$ is the log dividend yield. Assuming the log dividend yield itself follows a

[6] In particular, $k \equiv -\log(\rho) - (1-\rho)\log(1/\rho - 1)$ and $\rho \equiv 1/(1 + \exp(\overline{d-p}))$.

first-order autoregression

$$x_{t+1} = \phi x_t + u_{t+1}$$

and substituting this into the previous expression, we have $r_{t+1} \approx k + (1-\rho\phi)x_t + \varepsilon_{t+1}$, where $\varepsilon_{t+1} \equiv \Delta d_{t+1} - \rho u_{t+1}$ is a stationary variable. In other words, $E_t r_{t+1} \approx k + \beta_1 x_t$, where $\beta_1 = (1-\rho\phi)$, and the subscript under the β signifies that we are using one-period returns. Using similar calculations, we can show that the k-period continuously compounded return, r^k_{t+1}, defined as $r^k_{t+1} = \sum_{i=0}^{k} r_{t+1+i}$, can be written as

$$r^k_{t+1} \approx \tilde{k} + \left[(1-\rho\phi)\sum_{i=0}^{k-1} \phi^i\right] x_t + \tilde{\varepsilon}_{t+1} \approx \tilde{k} + \beta_k x_t + \tilde{\varepsilon}_{t+1} \quad (3)$$

where $\beta_k = (1-\rho\phi)\sum_{i=0}^{k-1} \phi^i$. This is the long-horizon regression in Fama and French (1988) and Campbell et al. (1997), and corresponds to my case 1. The additional insight is that I have used the log-linearized model to impose the restriction $\beta_k = (1-\rho\phi)\sum_{i=0}^{k-1} \phi^i$, which links β_k and ϕ. Given that $k = [\lambda T]$, and ρ is close to 1,

$$\lim_{T\to\infty} \beta_{[\lambda T]} = 1 - \mathrm{e}^{c\lambda}.$$

Since we can estimate β_k consistently, we can also find a consistent estimate of c by using the transformation $\hat{c}^{\text{CONSISTENT}} = \log(1-\hat{\beta}_k)/\lambda$, where $\hat{\beta}_k$ is the least squares estimate in (3). Moreover, we can use the delta method to find the asymptotic distribution of $c^{\text{CONSISTENT}}$. The estimator is super-consistent converging at rate T, and its asymptotic distribution is a mixture of normals (see Valkanov, 1998, for more details). The central idea in the above estimation procedure is to use information provided by a model in order to estimate c.

References

Barberis, N., 2000. Investing for the long run when returns are predictable. Journal of Finance 55, 225–264.

Boudoukh, J., Richardson, M., 1993. Stock returns and inflation: a long-horizon perspective. American Economic Review 64, 1346–1355.

Brennan, M., Schwartz, E., Lagnado, R., 1997. Strategic asset allocation. Journal of Economic Dynamics and Control 21, 1377–1403.

Campbell, J., 2001. Why long horizons? A study of power against persistent alternatives. Journal of Empirical Finance 8, 459–491.

Campbell, J., Shiller, R., 1987. Cointegration and tests of present value models. Journal of Political Economy 95, 1062–1088.

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

Campbell, J., Viceira, L., 1999. Consumption and portfolio decisions when expected returns are time varying. Quarterly Journal of Economics 114, 433–495.

Campbell, J., Lo, A., MacKinlay, A., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton.

Cavanagh, C., Elliott, G., Stock, J., 1995. Inference in models with nearly integrated regressors. Econometric Theory 11, 1131–1147.
Cochrane, J., 1999. New Facts in Finance, Unpublished working paper, University of Chicago.
Fama, E., French, K., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.
Ferson, W., Sarkissian, S., Simin, T., 2003. Spurious regressions in financial economics? Journal of Finance, forthcoming.
Fisher, M., Seater, J., 1993. Long-run neutrality and superneutrality in an ARIMA framework. American Economic Review 83, 402–415.
Goetzmann, W., Jorion, P., 1993. Testing the predictive power of dividend yields. Journal of Finance 48, 663–679.
Granger, C., Newbold, P., 1974. Spurious regressions in econometrics. Journal of Econometrics 2, 111–120.
Hansen, B., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 485–500.
Hansen, L., Hodrick, R., 1980. Forward exchange rates as optimal predictors of future spot rates. Journal of Political Economy 88, 829–853.
Hodrick, R., 1992. Dividend yields and expected stock returns: alternative procedures for inference and measurement. Review of Financial Studies 5, 357–386.
Lanne, M., 2002. Testing the predictability of stocks returns. Review of Economics and Statistics 84, 407–415.
Liu, J., 1999. Portfolio selection in stochastic environments. Unpublished working paper, University of California, Los Angeles.
Mishkin, F., 1990. What does the term structure of interest rates tell us about future inflation? Journal of Monetary Economics 70, 1064–1072.
Mishkin, F., 1992. Is the Fisher effect for real? Journal of Monetary Economics 1992, 195–215.
Nelson, C., Kim, M., 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 semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703–708.
Ng, S., Perron, P., 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation. Journal of the American Statistical Association 90, 268–281.
Phillips, P., 1986. Understanding spurious regressions in econometrics. Journal of Econometrics 33, 311–340.
Phillips, P., 1987. Toward a unified asymptotic theory of autoregression. Biometrika 74, 535–547.
Phillips, P., 1991. Optimal inference in cointegrated series. Econometrica 59, 283–300.
Richardson, M., Stock, J., 1989. Drawing inferences from statistics based on multi-year asset returns. Journal of Financial Economics 25, 323–348.
Schwert, W., 1987. Effects of model specification on tests for unit roots in macroeconomic data. Journal of Monetary Economics 20, 73–103.
Schwert, W., 1989. Tests for unit roots: a Monte Carlo investigation. Journal of Business and Economic Statistics 7, 147–159.
Schwert, W., 1990. Indexes of the United States stock prices from 1802 to 1987. Journal of Business 63, 399–426.
Siegel, J., 1992. The real rate of interest from 1800–1990: a study of the U.S. and U.K. Journal of Monetary Economics 29, 227–252.
Stambaugh, R., 1986. Bias in regressions with lagged stochastic regressors. Unpublished working paper, University of Chicago.
Stambaugh, R., 1999. Predictive regressions. Journal of Financial Economics 54, 375–421.
Stock, J., 1991. Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series. Journal of Monetary Economics 28, 435–459.
Stock, J., 1994, Unit roots, structural breaks and trends. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. IV, Elsevier, Amsterdam, pp. 2740–2841.

Stock, J., Watson, M., 1989. Interpreting the evidence on money-income causation. Journal of Econometrics 40, 161–182.

Torous, W., Valkanov, R., 2002. Boundaries of predictability: noisy predictive regressions. Unpublished working paper, University of California, Los Angeles.

Torous, W., Valkanov, R., Yan, S., 2002. On predicting stock returns with nearly integrated explanatory variables. Journal of Business, forthcoming.

Valkanov, R., 1998. The term structure with highly persistent interest rates. Unpublished working paper, University of California, Los Angeles.

Viceira, L., 1997. Testing for structural change in the predictability of asset returns. Unpublished working paper, Harvard University.

Watson, M., 1994. Vector autoregressions and cointegration. In: Engle, R., McFadden, D. (Eds.), Handbook of Econometrics, Vol. IV, Elsevier, Amsterdam, pp. 2844–2915.