Long-Horizon Regressions: Theoretical Results and Applications to the Expected Returns/Dividend Yields and Fisher Effect Relations

Author
Valkanov, Rossen

Publication Date
1999-09-09

Long-Horizon Regressions: Theoretical Results and Applications to the Expected Returns/Dividend Yields and Fisher Effect Relations[1]

Rossen Valkanov[2]

09/09/1999

[1] I have benefitted from discusions with Ivo Welch and Pedro Santa-Clara and from the detailed comments of Mark Watson and Yacine Aït-Sahalia. All remaining errors are my own.

[2] Contact: rossen.valkanov@anderson.ucla.edu or Anderson Graduate School of Management at UCLA, 110 Westwood Plaza, Box 951481, Los Angeles, CA 90095-1481, tel: (310) 825-7246.

Abstract

We analyze several ways of conducting long-horizon regressions, taken from the empirical literature. Asymptotic arguments are used to show that, in all cases, the t-statistics do not converge to well-defined distributions, thus explaining the tendency of long-horizon regressions to find 'significant' results, where previous short-term approaches have failed. Moreover, in some cases, the ordinary least squares estimator is not consistent, and the R^2 cannot be interpreted as a measure of the goodness of fit. Those results cast doubt on the conclusions reached by most previous long-horizon regression studies. We propose a rescaled t-statistic, whose asymptotic distribution is easy to simulate, and re-visit some of the evidence on the long-horizon predictability of returns and the long-horizon tests of the Fisher Effect.

1 Introduction

There has been an increasing interest in long-horizon regressions, since 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 (1988), Mishkin (1990, 1992), Boudoukh and Richardson (1993), 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 a long-horizon variable is obtained as a rolling sum of the original series. 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 expected returns/dividend yields, the Fisher Effect, or neutrality of money, the striking results produced by those studies prompted us to scrutinize the appropriateness of the econometric methods.

In this paper, we show that long-horizon regressions will always produce 'significant' results, whether or not there is a relationship 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 1 (or I(1)). Such a persistent stochastic behavior will be observed whenever the regressor, the regressand, or both, are obtained by summing over a non-trivial fraction of the sample. Based on this insight, we use the Functional Central Limit Theorem to analyze the distributions of statistics from long-run regressions, commonly used in economics and finance. We 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 1 in probability. Those results remind us, but are not analogous to the ones in Granger and Newbold (1974), and explained by Phillips (1986). The analogy comes from 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. Richardson and Stock (1989) use a methodology similar to our, but they consider only univariate regressions. Their results can be viewed as a special case in our framework.

We provide a simple guide on how to conduct estimation and inference using long-horizon regressions. The focus is on the asymptotic properties of the OLS estimator of the slope coefficient, its t-statistic and the coefficient of determination. Various ways of conducting long-horizon regressions are analyzed, all taken from previous studies. The estimators from some regressions, frequently used in empirical work, are inconsistent. Moreover, the t-statistics from all the considered regressions do not converge to well-defined distributions, thus putting into question the conclusions from studies that use long-run variables.

We propose a rescaled t-statistic, $t/\sqrt{T}$, for testing long-horizon regressions. Its asymptotic distribution, although non-normal, is easy to simulate. We use the proposed methods to re-examine the expected returns/dividend yield regressions in Fama and French (1988), and the Fisher Effect tests in Boudoukh and Richardson (1993), and Mishkin (1992). The obtained results are quite general and applicable whenever long-horizon regressions are employed.

The analytical expressions that we provide explain some of the empirical and simulation results obtained by previous authors. For example, we can tackle the interesting question of whether long-horizon regressions have greater power to detect deviations from the null than do short-horizon regressions, or are the significant results a mere product of size distortion. This question, indirectly addressed in Hodrick (1992), Mishkin (1990, 1992), Goetzmann and Jorion (1993) and Campbell (1993), has been posed explicitly in Campbell, Lo, and MacKinlay (1997). Some Monte-Carlo simulations suggest power gains (Hodrick (1992)), other show size distortions (Goetzmann and Jorion (1993)), but a definite, analytic answer has not yet been provided. Our results show that the significant results from long-horizon regressions are due to incorrect critical values. Another implication of our analysis is that a significant R^2 in such regressions cannot be interpreted as an indication of a good fit.

This paper is not a condemnation of long-horizon studies. Our 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, our analysis provides a general guideline on how to test long-horizon relations. Some researchers are aware that normal asymptotic approximations are not adequate when using long-horizon variables. However, the reason for the poor approximations is attributed to serial correlation, induced in the error terms while transforming the data, and to endogeneity bias. Serial correlation will lead to consistent estimates, but for testing, the standard errors are produced with the Hansen-Hodrick (1980) or Newey-West (1987) method.

Even after correcting for serially correlated errors, Monte Carlo simulations show that 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)). We discuss the reasons for the poor approximation and suggest alternative methods for testing and estimation in long-horizon regressions.

The paper is structured as follows. Section 2 presents the various ways of specifying long-horizon regressions that have been commonly used in the empirical literature. These regressions can be categorized into four cases, arising naturally from the choice of a null hypothesis and the model being considered. Section 3 provides the main theoretical results. Testing and inference in all four cases is analyzed, using asymptotic methods. In section 4, we conduct simulations to illustrate the analytical results. Section 5 applies the conclusions in section 3 to the expected 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 Models

The underlying data-generating processes are:

$$\begin{aligned} Y_{t+1} &= \alpha + \beta X_t + \varepsilon_{1,t+1} && (1) \\ (1-\phi L)\, b(L) X_{t+1} &= \mu + \varepsilon_{2,t+1} && (2) \end{aligned}$$

The variable X_t is represented as an autoregressive process, whose highest root, ϕ, is conveniently factored out[1] and $b(L) = b_0 + b_1 L + b_2 L^2 + ... + b_p L^p$ is invertible. For simplicity, we let $\phi = 1$, although all that is required is a

[1] See Stock (1991).

highly persistent process[2]. Define $w_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, where w_t is a martingale difference sequence with $E\left(w_t w_t' | w_{t-1}, ...\right) = \Sigma = [\sigma_{11}^2 \;\; \sigma_{12}; \sigma_{21} \;\; \sigma_{22}^2]$ and finite fourth moments.

We could have started with a general vector autoregression (VAR) (c.f 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 expected return or the equity premium. In the inflation/interest rate literature, X_t denotes the interest rate. Similarly, in the money neutrality literature, X_t represents the nominal money supply. The timing of the variables is chosen to conform exactly with the previous literature on expected returns and dividend yields (Stambaugh (1986, 1999), Cavanagh et al (1994)), but it is not essential for the results. We could well have started with a more general triangular cointegrated system (Campbell and Shiller (1987), Phillips (1991))[3]. 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.

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[4]. 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

[2] This assumption can be relaxed, by letting ϕ be in the neighborhood of one (c.f. Phillips (1987), Stock (1991), Cavanagh et al. (1994), Torous (1999?))

[3] Note that we can always express a triangular cointegrated system with (1-2), whereas (1-2) cannont necessarily be represented as a triangular cointegrated system. However, the distinction is insignificant in practical applications.

[4] This is not surprising in the equity premium literature, since there is much more noise in equation (1) than signal coming from X_t, as discussed in Valkanov (1999)

to strengthen the signal, while eliminating the noise. Given $(1-2)$, the long-horizon variables are:

$$\begin{aligned} Z_t^k &= \sum_{i=0}^{k-1} Y_{t+i} \\ Q_t^k &= \sum_{i=0}^{k-1} X_{t+i} \end{aligned}$$

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). Example of the first type are the papers by Fama and French (1988) and Campbell and Shiller (1988). In the equity/dividend literature, Z_t^k is the k-th period continuously compounded (log) return. Examples of the second type of long-horizon regressions are the papers by Boudoukh and Richardson (1993), Mishkin (1992), and Fisher and Seater (1993), where Z_t^k is the k-th period continuously compounded (log) return and the k-th period GDP growth, respectively, and Q_t^k is the k-th 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 equity premium literature, it is appropriate to assume that dividends have no predictive power for expected returns, or $\beta = 0$ (Fama and French (1988) and 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) and Boudoukh and Richardson (1993)). In the money neutrality literature, tests are usually carried out under the null of $\beta = 0$ (Stock and Watson (1989) and Fisher and Seater (1993)). Hence, a comprehensive analysis of long-horizon regressions based on (1-2) must accommodate the null hypotheses of $\beta = 0$ and $\beta = \beta_o \neq 0$.

The empirically interesting specifications of long-horizon regressions can be categorized into four cases, presented in Table 1. The regressand is always

Specification of long-horizon regressions

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

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

$Z_{t+1}^k = \sum_{i=0}^{k-1} Y_{t+1+i}$, but the regressor and the relationship under the null vary. In the first two 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 $\beta = \beta_o \neq 0$. 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).

We 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 and by Valkanov (1998) in estimating the persistence of short-term interest rates. More specifically, Richardson and Stock (1989) 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 to specify long-horizon regressions. We prove that the statistics in long-horizon relations do not have the convenient properties, namely, consistent OLS estimators of the slope coefficient, t-tests with adequate power and size, and R^2 that converge to 1 in probability under the null. More importantly, in *all* cases, the t-statistic fails to converge to a well-defined distribution, but explodes as the sample size increases. In finite samples, this results in increasingly significant t-tests, as the overlap increases, whether or not there is a relationship 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 reasons underlying the results are simple. Aggregating a fraction of the sample produces a persistent variable that behaves very much like an I(1) process. The time-dependence is strong enough that correcting for serial correlation using the Hansen-Hodrick-Newey-West procedure is not sufficient. 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, we present the analytical results. In addition to stating the theorems, we also provide informal discussions to clarify their implications and applications.

The assumptions and additional notation are summarized here for convenience.

Assumptions 1 *In model* $(1-2)$, *let*

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. $\phi = 1$ *and* $\alpha = \mu = 0$

4. $w_t = (\varepsilon_{1,t}, \varepsilon_{2,t})'$, *where* w_t *is a martingale difference sequence with* $E\left(w_t w_t' | w_{t-1}, ...\right) = \Sigma = [\sigma_{11}^2 \ \ \sigma_{12}; \sigma_{21} \ \ \sigma_{22}^2]$, *and finite fourth moments.*

5. *The roots of* $b(L) = b_0 + b_1 L + b_2 L^2 + ... + b_p L^p$ *are less than* 1 *in absolute value,* $\sum_{i=1}^{p} i\,|b_i| < \infty$, *and* p *is a fixed number.*

It is known that under the above assumptions, $\left(\frac{1}{\sqrt{T}\sigma_{11}} \sum_{i=0}^{[sT]} \varepsilon_{1,i}, \frac{1}{\sqrt{T}\sigma_{22}} \sum_{i=0}^{[sT]} \varepsilon_{2,i}\right) \Rightarrow (W_1(s), W_2(s))$ jointly, from the Functional Central Limit Theorem (FCLT), where $\Rightarrow$ denotes weak convergence, and $W_1(s)$ and $W_2(s)$ are two standard Weiner processes on $D[0,1]$, with covariance $\delta = \sigma_{12}/(\sigma_{11}\sigma_{22})$. Similarly, if $\omega^2 = \sigma_{11}^2/b^2(1)$, then $\frac{1}{\omega\sqrt{T}} X_{[sT]} \Rightarrow W_2(s)$.

Assumptions 1 simplify the analysis but can easily be generalized. For instance, ϕ can be in a neighborhood of one, instead of exactly at unity (see Phillips (1987), Stock (1991), and Cavanagh et al. (1994)). We may allow w_t to follow a less restrictive time series process as in Stock and Watson (1993). More generally, the results below will be valid for any strongly mixing process w_t, thus allowing for weakly dependent and possibly heterogeneous innovations (Hansen (1992)). All regressions are run with a constant, as it is usually done in practice.

Before proceeding with the main results, we prove a lemma that will serve as a foundation for the rest of the paper. In this lemma, we show that long-horizon variables, which are nothing but partial sums of the underlying processes X_t and Y_t, converge weakly to functionals of diffusion processes, after appropriate rescaling.

- **Lemma 1** *If assumptions 1 hold, then*

  1. $\frac{1}{\omega T^{3/2}} Q_t^k \Rightarrow \int_s^{s+\lambda} W_2(\tau) d\tau \equiv \overline{W}_2(s;\lambda)$
  2. $\frac{1}{\omega T^{3/2}} \left(Q_t^k - \overline{Q}^k\right) \Rightarrow \overline{W}_2(s;\lambda) - \frac{1}{1-\lambda} \int_0^{1-\lambda} \overline{W}_2(s;\lambda) ds \equiv \overline{W}_2^{\mu}(s;\lambda)$

  If $\beta = 0$, *then*

3. $\frac{1}{\sqrt{T}\sigma_{11}} Z_t^k \Rightarrow W_1(s+\lambda) - W_1(s) = W_1(s;\lambda)$

4. $\frac{1}{\sqrt{T}\sigma_{11}} \left(Z_t^k - \overline{Z}^k\right) \Rightarrow W_1(s;\lambda) - \frac{1}{1-\lambda}\int_0^{1-\lambda} W_1(s;\lambda)ds = W_1^{\mu}(s;\lambda)$

   and if $\beta \neq 0$, *then*

5. $\frac{1}{\omega T^{3/2}} Z_t^k \Rightarrow \beta \overline{W}_2(s;\lambda)$

6. $\frac{1}{\omega T^{3/2}} \left(Z_t^k - \overline{Z}^k\right) \Rightarrow \beta \overline{W}_2^{\mu}(s;\lambda)$

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

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

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

$$F_3(A(s),B(s)) \equiv \frac{\left(\int_0^{1-\lambda}(A(s)B(s))\,ds\right)^2}{\int_0^{1-\lambda}(A(s))^2 ds \int_0^{1-\lambda}(B(s))^2 ds}$$

All distributions can be represented as one of those 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 hold, if we regress* Z_t^k *on a constant and* X_t*, the slope coefficient and the associated statistics will have the following properties:*

- $\hat{\beta} \Rightarrow \frac{\sigma_{11}}{\omega} F_1\left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s)\right)$
- $\frac{t_{\hat{\beta}}}{T^{1/2}} \Rightarrow F_2\left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s)\right)$
- $R^2 \Rightarrow F_3\left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s)\right)$

Some results are worth emphasizing. First, $\hat{\beta}$ is not a consistent estimator of β and its distribution depends explicitly on σ_{11} and ω, which can be estimated consistently from $(1-2)$. Second, the t-statistic, testing for $\beta = 0$, 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 by carrying 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 parameter δ, which can be estimated consistently from $(1-2)$. Lastly, the coefficient of determination, R^2, does not converge to 1 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 hold, if we regress Z_t^k on a constant and Q_t^k, the slope coefficient and the associated statistics will have the following properties:*

- $T(\hat{\beta} - 0) \Rightarrow \frac{\sigma_{11}}{\omega} F_1 \left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda) \right)$
- $\frac{t}{T^{1/2}} \Rightarrow F_2 \left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda) \right)$
- $R^2 \Rightarrow F_3 \left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda) \right)$

This is case 2. Here $\hat{\beta}$ is a consistent estimator of β. However, the t-statistic does not converge to a well-defined distribution. Similarly to case 1, the R^2 does not converge in probability to 1. 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 hold, if we regress Z_t^k on a constant and X_t, the slope coefficient and the associated statistics will have the following properties:*

- $\frac{\hat{\beta}}{T} \Rightarrow \beta F_1 \left(W_2^{\mu}(s), \overline{W}_2^{\mu}(s;\lambda) \right)$
- $\frac{t_{\hat{\beta}}}{T^{1/2}} \Rightarrow F_2 \left(W_2^{\mu}(s), \overline{W}_2^{\mu}(s;\lambda) \right)$
- $R^2 \Rightarrow F_3 \left(W_2^{\mu}(s), \overline{W}_2^{\mu}(s;\lambda) \right)$

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 sample size or the overlap increases, one would tend to observe increasing (in magnitude) slope coefficients, as is the case in Fama and French (1998), Campbell and Shiller (1988), and Boudoukh and Richardson (1993). Moreover, the limiting distribution of $\frac{\hat{\beta}}{T}$ depends on the unknown parameter β itself. Similarly 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. Fortunately, the limiting distribution does not depend on any unknown parameters and can easily be simulated. The coefficient of determination does not converge to 1 in probability, but has a well defined distribution. However, it cannot be used to judge the fit of the regression in the usual fashion. In case 3, the limiting distributions are invariant to δ.

Lastly, we present the results for case 4.

Theorem 5 *When $\beta \neq 0$ and assumptions 1 hold, if we regress Z_t^k on a constant and Q_t^k, the slope coefficient and the associated statistics will have the following properties:*

- $T(\hat{\beta} - \beta) \Rightarrow \frac{\sigma_{11}}{\omega} F_1\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right)$
- $\frac{t_{\hat{\beta}}}{T^{1/2}} \Rightarrow F_2\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right)$
- $R^2 \rightarrow^p 1$

In many respects, this is the econometrically 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 a consistent estimate of 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 relationship between Y_t 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 exactly the same for cases 2 and 4.

Table 2 provides a quick summary of the above results[5]. In all four cases, inference based on the t-statistic cannot be done using asymptotic critical values. However, the normalized t-statistic, $t/\sqrt{T}$, has a well-defined distribution that can be easily simulated. Moreover, as the Monte-Carlo 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

[5] The stochastic order notation $V_T = O_p(1)$ intuitively means that the variable V_T has a well defined distribution, as $T \to \infty$.

Specification of long-horizon regressions

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:	$T(\hat{\beta} - \beta) = O_p(1)$ $\frac{t}{T^{1/2}} = O_p(1)$ $R^2 \to^p 1$

Table 2: Summary of the results in Theorems 2-5.

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 1 only in theorem 5. In the other cases, it can lead to false conclusions as demonstrated in the next section.

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 very fast to their limiting distributions (Stock (1994)), or in other words, the asymptotic distributions can be approximated well with a relatively small sample size. Here, we conduct Monte-Carlo simulations to illustrate some of the points made in the previous section. First, we demonstrate that the rescaled t-statistic does converge asymptotically for all of the above cases, unlike the non-rescaled t-statistic. Second, we plot the densities, derived in Theorems 2-5. Third, we investigate how those densities depend on δ, the correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$.

The experiment is conducted as follows. For each case, we simulate series of length 100 and 750 observations 5000 times. The first simulation corresponds to a typical annual data set, whereas the second corresponds to a series with monthly frequencies. For each simulation, we compute $\hat{\beta}$, the t-statistic under the appropriate null hypothesis, and R^2. For simplicity, we let $\sigma_{11}^2 = \omega^2 = 1$ without losing generality, since the asymptotic distributions of $t/\sqrt{T}$ and R^2 are invariant to those parameter. The densities of the simulations for each case are plotted in figures 1 through 4. In each figure, the first column of graphs displays the densities of the non-normalized statistics, whereas the second column displays those of the appropriately rescaled statistics, as suggested by the theorems above. In the cases where no rescaling is needed for convergence, the graphs across columns are identical.

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 that can easily be simulated, provided we have an estimate of δ. The nuisance parameter, δ can be estimated consistently from $(1-2)$. In the first set of simulations, we let $\delta = 0$ and approximate the limiting distributions by rescaled sums of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$. The results are presented in the second column of figures 1-4. Notice that the partial sums converge very fast to their limiting distributions; the estimated densities for $T = 100$ and $T = 750$ are almost identical.

The R^2 converges to 1 only in case 4. In cases 1 and 2, its density has most of its mass under 0.5. In case 3, most of the mass is between 0.1 and 0.98. Finally, the figures show that $\hat{\beta}$ is not consistent in cases 1 and 3; its variance does not decrease as T increases. Table 3a provides the first two moments of all the statistics in cases 1-4, to confirm the above conclusions.

The previous results were obtained under the assumption of no correlation between $W_1(s)$ and $W_2(s)$, or $\delta = 0$. Now, we repeat the same battery of simulations for $\delta = 0.9$. To investigate the effect of this change, we plot the limiting distributions of $t/\sqrt{T}$ for $\delta = 0$ together with those from $\delta = 0.9$, in figure 5. In case 1, the mean of the distribution changes. For cases 2 and 4, the limiting distributions are identical, as discussed above. However, an increase in the correlation δ alters the shape of the distribution. In case 3, the asymptotic distribution of $t/\sqrt{T}$ is invariant to δ, as mentioned above. Table 3b provides the first two moments of all the statistics, in cases 1-4, for $\delta = 0.9$.

5 Empirical Results–The Expected Returns/Dividend Yield and Fisher-Effect regressions

5.1 Long-Horizon Predictability of Expected Returns using

Dividend Yields

The predictability of expected 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 (Barberis (1998), Brennan, Schwartz, and Lagnado (1997), Campbell and Viceira (1998), Campbell and Cochrane (1999), and Liu (1999), among others) that try to analyze the implications of return predictability on 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-test and R^2. Both of those statistics increase with the horizon. However, Goetzmann and Jorion (1993) conduct simulations and bootstrap studies on the properties of the estimates from the equity premium/dividend yield long-horizon regressions and conclude that the evidence for predictability is not nearly as overwhelming at long-horizons as the previous studies have suggested. Hodrick (1992) and Kim and Nelson (1993) use VAR simulation-based studies conclude that the distortions in the distributions of the t-statistics were not enough to overturn the conclusion of Fama and French (1988).

It is difficult to reconcile or compare the results from these studies, because they are all based on simulations or bootstrap, and fail to yield general results. Their conclusions are ultimately a function on how the artificial data is being generated. More importantly there is no way of knowing what features of the data would influence the statistics of interest and in what fashion.

In section 3, we provided a framework which will now be applied to explain and re-test the results from Fama and French (1988). Fortunately enough, re-testing is simple and does not require a replication of the regressions from the original study. The results from the relevant t-statistics, reported in Fama and French are normalized by the square root of the sample size, yielding a new statistic, which, unlike the previous one, has a well-defined asymptotic distribution. The distribution under the null of no predictability, derived in theorem 2 (case 1), can be simulated using a consistent estimator of δ. We estimate δ from the residuals of regressions $(1-2)$ using CRSP data from 1927 to 1986. The lag structure of the dividend yield series is chosen with sequential t-tests, as suggested by Ng and Perron (1995, 1998). The selected lag structure has 14 lags, yielding a highest autoregressive root of 0.96. Note that previous studies have arbitrarily chosen the dividend yield to follow an AR(1) process and the estimated highest autoregressive root is lower, 0.85 (Nelson and Kim (1993)). We obtain an estimate of $\hat{\delta} = -0.04$. In other words, after capturing the dynamics of the dividend yield, the residuals of the two regressions are not very correlated, contrary to the results reported in Nelson and Kim (1993) and Stambaugh (1999)[6].

Table 4 presents the results from table 3 in Fama and French (1988) for nominal returns, using dividend yields and the dividend price ratio as regressors. The conclusions for real returns are identical and, hence, omitted. Along with the estimates $\hat{\beta}$ and the t-statistics, we compute the $t/\sqrt{T}$ statistic. Given that δ is consistently estimated, we simulate the distribution from Theorem 2. The percentile of the normalized t-statistic, reported in table 4, allows us to test the null $\beta = 0$ at various levels of significance. We would reject if the t-statistic falls in the tails of the distribution. For example, values lower than 0.05 or higher than 0.95 indicate rejection at the 10% level, values lower than 0.025 or higher than 0.975 indicate rejection at the 5% level, and so forth.

[6]Note that δ does not correspond exactly to the correlation between the error terms in Stambaugh (1986, 1999) and Nelson and Kim (1993).

Using the entire sample of 1927-1986, we cannot reject the null of no predictability at conventional levels of significance. Looking at the sub-periods, the lack of rejection seems to come from the 1927-1956 period. Indeed, after 1956, there seems to be more evidence of predictability. In the post-1956 periods, the null can be rejected at the 5% level using the dividend-price ratio as predictor, but not at the 1% level. The dividend yield performs worse than the dividend price ratio in all periods and for all maturities. To sum up the results from table 4, the $t/\sqrt{T}$ statistic suggests that there is little evidence of returns predictability before 1956. In the 1956-1986 period, the dividend-price ratio and the dividend yield seem to have some predictive power. Overall, the dividend-price ratio is more successful in explaining variations in expected returns.

5.2 Long-Horizon Fisher Effect

The Fisher Effect, positing that nominal returns must equal real returns plus expected inflation, has been estimated and tested by many authors in different forms. We 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). Both of those papers cannot reject the null that long-horizons 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? We revisit the main results from both papers below.

5.2.1 Nominal Stock Returns and Inflation

Boudoukh and Richardson (1993) 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. The authors are concerned about measurement problems that might arise from using ex-post instead of ex-ante inflation. To remedy the problem, they also estimate the regression above using several instruments, one of which is lagged long-horizon inflation, $\sum_{i=1}^{5} \pi_{t-5+i}$. We revisit the OLS and the IV calculations using the estimates from Boudoukh and Richardson, tables 1 and 2. The normalized t-statistics is computed by dividing the reported t-statistic by the square root of the sample size. The results are presented in table 3 below for the entire sample and two sub-samples.

We obtain an estimate of $\hat{\delta} = 0.12$ using the data from Siegel (1992) and Schwert (1990)[7] and simulate the limiting distribution from theorem 5. 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 distributions, because of the lags. The results from theorem 4 for the IV case are: $T(\hat{\beta} - \beta) \Rightarrow \frac{\sigma_{11}}{\omega} F_1 \left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s-\lambda; s) \right)$, $\frac{t_{\hat{\beta}}}{T^{1/2}} \Rightarrow F_2 \left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s-\lambda; s) \right)$, $R^2 \to^p 1$.

Rejection would occur if the statistic falls in the tails of the distribution. Values lower than 0.05 or higher than 0.95 indicate rejection at the 10% level. For the OLS case, the null of a full Fisher Effect cannot be rejected at the 5% level, except for the 1914-1990 sub-period, where it cannot be rejected at

[7] The estimates of δ for the sub-samples were very close to the one from the entire sample. Since we have no reasons to suspect variations in this parameter, we use the more precise estimate from the 1802-1990 period.

the 1% level. This conclusion contradicts the one from the usual t-statistics, where rejection occurs for all periods. In the IV case, the null cannot be rejected for any period. We also performed tests under the null $\beta = 0$, corresponding to case 1. The null was rejected at all levels of significance, thus providing indirect indication for good testing power. To conclude, our re-examination of the evidence by Boudoukh and Richardson (1993) supports a full Fisher Effect. The evidence is particularly strong in the first sub-period.

5.2.2 Inflation and Nominal Interest Rates

Mishkin (1992) focuses on a similar regression using monthly macroeconomic data[8]. The regressor is the 3-period interest rate, whereas and the regressand is the 3-period inflation[9]. Despite the relatively small overlap in the creation of the new series, Monte Carlo simulations conducted by the author (tables 3-4) convinced us that our 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 sub-sample. Recalling figure 4, 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 from creating long-horizon variables. Indeed $\beta = 1$ is rejected using conventional critical values (table 1, Mishkin (1992)).

[8] The data is descibed in Mishkin (1990)

[9] Let $p_t = \log(P_t)$, where P_t is the price level at time t. The 3-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, $\left(1 - L^k\right) Y_{t+k} = (1 - L)\left(1 + L + ... + L^{k-1}\right) Y_{t+k} = \sum_{i=1}^{k} \triangle Y_{t+i}$

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 table 6, for the period 1953-1990 and different sub-periods. The last column shows the percentile of the computed $t/\sqrt{T}$ statistic under the null. 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). However, the null cannot be rejected at the 5% level (two-sided test) for any period by using the $t/\sqrt{T}$ statistic. The evidence of 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 (table 4), even though he does not provide an econometric explanation to account for the results.

The $t/\sqrt{T}$ tests of long-horizon Fisher Effect in the stock (Boudoukh and Richardson (1993)) and bond (Mishkin (1992)) market 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. Further testing and exploration of this apparent structural change might be an interesting question for future research.

6 Conclusion

We 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. We 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. to run 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 sample size or 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, when it is true. In order to conduct asymptotically valid tests, we 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 to 1 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.

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 is, 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. In the last section of this paper, the proposed $t/\sqrt{T}$ statistic is employed to re-examine the predictability of returns in Fama and French (1988) using dividend yields and the dividend-price ratio. We find little predictability for the periods before 1956. In the 1956-1986 period, the dividend-price ratio and the dividend yield seem to have some predictive power. In another application, the $t/\sqrt{T}$ statistic is used to 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

Proof of Lemma 1: Recall that $\frac{1}{\omega T^{1/2}} X_t \Rightarrow W_2(s)$, where from now onward $t = [sT]$ 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 $\frac{1}{\omega T^{3/2}} Q_t^k = \frac{1}{T} \sum_{i=0}^{k-1} \frac{X_{t+i}}{\omega T^{1/2}} \Rightarrow \int_s^{s+\lambda} W_2(\tau) d\tau \equiv \overline{W}_2(s;\lambda)$, using the continuous mapping theorem (CMT) to prove part 1. Similarly, for part 2, $\frac{1}{\omega T^{3/2}} \overline{Q}^k = \frac{1}{T-k} \sum_{t=1}^{T-k+1} \frac{1}{\omega T^{3/2}} Q_t^k \Rightarrow \frac{1}{1-\lambda} \int_0^{1-\lambda} \overline{W}_2(s;\lambda) ds$, using the CMT. Then, we can write, $\frac{1}{\omega T^{3/2}} \left(Q_t^k - \overline{Q}^k \right) \Rightarrow \overline{W}_2(s;\lambda) - \frac{1}{1-\lambda} \int_0^{1-\lambda} \overline{W}_2(s;\lambda) ds \equiv \overline{W}_2^{\mu}(s;\lambda)$.

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

When $\beta \neq 0, Y_t = \beta X_{t-1} + \varepsilon_{1,t}$. Therefore, $\frac{1}{\omega T^{3/2}} Z_t^k = \frac{1}{\omega T^{3/2}} \sum_{i=0}^{k-1} Y_{t+i+1} = \beta \frac{1}{\omega T^{3/2}} \sum_{i=0}^{k-1} X_{t+i} + \frac{1}{\omega T^{3/2}} \sum_{i=0}^{k-1} \varepsilon_{1,t+i} \Rightarrow \beta \overline{W}_2(s;\lambda)$ using part 1 and $\frac{1}{\omega T^{3/2}} \sum_{i=0}^{k-1} \varepsilon_{1,t+i} = o_p(1)$. Part 6 is proven in exactly the same fashion. ■

Proof of Theorem 2: By definition, $\hat{\beta} = \frac{\sum_{t=1}^{T-k} \left(Z_{t+1}^k - \overline{Z}^k \right) \left(X_t - \overline{X} \right)}{\sum_{t=1}^{T-k} \left(X_t - \overline{X} \right)^2} =$

$$\frac{\frac{1}{T} \sum_{t=1}^{T-k} \left(\frac{Z_{t+1}^k - \overline{Z}^k}{T^{1/2}} \right) \left(\frac{X_t - \overline{X}}{T^{1/2}} \right)}{\frac{1}{T} \sum_{t=1}^{T-k} \left(\frac{X_t - \overline{X}}{T^{1/2}} \right)^2} \Rightarrow \frac{\sigma_{11}}{\omega} \frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda) W_2^{\mu}(s) ds}{\int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds} = \frac{\sigma_{11}}{\omega} F_1 \left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s) \right)$$

using Lemma 1, part 4 and the CMT. Define $s^2 = \frac{1}{T} \sum_{t=1}^{T-k} \left(Z_{t+1} - \overline{Z}^k \right)^2 - \hat{\beta}^2 \frac{1}{T} \sum_{t=1}^{T-k} \left(X_t - \overline{X} \right)^2$. Recall that $\frac{1}{T} \sum_{t=1}^{T-k} \left(\frac{Z_{t+1} - \overline{Z}^k}{T^{1/2}} \right)^2 \Rightarrow \sigma_{11}^2 \int_0^{1-\lambda} \left(W_1^{\mu}(s;\lambda) \right)^2 ds$ and $\frac{1}{T} \sum_{t=1}^{T-k} \left(\frac{X_t - \overline{X}}{T^{1/2}} \right)^2 \Rightarrow \omega^2 \int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds$. Then, $\frac{s^2}{T} \Rightarrow \sigma_{11}^2 \int_0^{1-\lambda} \left(W_1^{\mu}(s;\lambda) \right)^2 ds - \frac{\sigma_{11}^2}{\omega^2} \left(\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda) W_2^{\mu}(s) ds}{\int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds} \right)^2 \omega^2 \int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds =$

$$\sigma_{11}^2 \left[\frac{\int_0^{1-\lambda} \left(W_1^{\mu}(s;\lambda) \right)^2 ds \int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds - \left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda) W_2^{\mu}(s) ds \right)^2}{\int_0^{1-\lambda} \left(W_2^{\mu}(s) \right)^2 ds} \right].$$ Under the null,

the t-statistic is: $t = \frac{(\hat{\beta}-0)\left(\sum_{t=1}^{T-k}(X_t-\overline{X})^2\right)^{1/2}}{s}$. In our case, $\frac{t}{T^{1/2}} = \frac{\hat{\beta}\left(\sum_{t=1}^{T-k}(X_t-\overline{X})^2\right)^{1/2}}{\left(\frac{s^2}{T}\right)^{1/2}} \Rightarrow$
$\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)W_2^{\mu}(s)ds}{\left[\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(W_2^{\mu}(s)\right)^2 ds - \left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)W_2^{\mu}(s)ds\right)^2\right]^{1/2}} = F_2\left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s)\right).$

The coefficient of determination is defined as: $R^2 = \hat{\beta}\frac{\sum_{t=1}^{T-k}(X_t-\overline{X})^2}{\sum_{t=1}^{T-k}\left(Z_{t+1}^k-\overline{Z}^k\right)^2} =$
$\hat{\beta}\frac{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{X_t-\overline{X}}{T^{1/2}}\right)^2}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}^k-\overline{Z}^k}{T^{1/2}}\right)^2} \Rightarrow \frac{\left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)W_2^{\mu}(s)ds\right)^2}{\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(W_2^{\mu}(s)\right)^2 ds} = F_3\left(W_1^{\mu}(s;\lambda), W_2^{\mu}(s)\right).$

■

Proof of Theorem 3: The OLS estimator is: $\hat{\beta} = \frac{\sum_{t=1}^{T-k}\left(Z_{t+1}^k-\overline{Z}^k\right)\left(Q_t^k-\overline{Q}^k\right)}{\sum_{t=1}^{T-k}\left(Q_t^k-\overline{Q}^k\right)^2}$.

Using the results in Lemma 1 and the CMT, $T\left(\hat{\beta}-0\right) = \frac{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}^k-\overline{Z}^k}{T^{1/2}}\right)\left(\frac{Q_t^k-\overline{Q}^k}{T2}\right)}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Q_t^k-\overline{Q}^k}{T^{3/2}}\right)^2} \Rightarrow$
$\frac{\sigma_{11}}{\omega}\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda)ds}{\int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds} = \frac{\sigma_{11}}{\omega}F_1\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right)$ as required. Similarly, $\frac{s^2}{T} = \frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}-\overline{Z}^k}{T^{1/2}}\right)^2 - \left(T\hat{\beta}\right)^2\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Q_t^k-\overline{Q}^k}{T^{3/2}}\right)^2 \Rightarrow$
$\sigma_{11}^2\left[\frac{\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds - \left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda)ds\right)^2}{\int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds}\right]$. The t-statistic under the null is: $t = \frac{(\hat{\beta}-0)\left(\sum_{t=1}^{T-k}\left(Q_t^k-\overline{Q}^k\right)^2\right)^{1/2}}{s}$, and $\frac{t}{T^{1/2}} = \frac{T(\hat{\beta}-0)\left(\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Q_t^k-\overline{Q}^k}{T^{3/2}}\right)^2\right)^{1/2}}{\left(\frac{s^2}{T}\right)^{1/2}} \Rightarrow$
$\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda)ds}{\left[\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds - \left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda)ds\right)^2\right]^{1/2}} = F_2\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right).$

In a similar fashion, $R^2 = \hat{\beta}\frac{\sum_{t=1t}^{T-k}\left(Q_t^k-\overline{Q}^k\right)^2}{\sum_{t=1}^{T-k}\left(Z_{t+1}^k-\overline{Z}^k\right)^2} = T\hat{\beta}\frac{\frac{1}{T}\sum_{t=1t}^{T-k}\left(\frac{Q_t^k-\overline{Q}^k}{T^{3/2}}\right)^2}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}^k-\overline{Z}^k}{T^{1/2}}\right)^2} \Rightarrow$
$\frac{\left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda)ds\right)^2}{\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds} = F_3\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right)$, completing the proof. ■

Proof of Theorem 4: The proofs follow exactly the same pattern.

The OLS estimator is: $\hat{\beta} = \frac{\sum_{t=1}^{T-k}\left(Z_{t+1}^k - \overline{Z}^k\right)\left(X_t - \overline{X}\right)}{\sum_{t=1}^{T-k}\left(X_t - \overline{X}\right)^2}$. However, since $\beta \neq 0$, we have $\frac{Z_{t+1}^k - \overline{Z}^k}{T^{3/2}} \Rightarrow \beta \overline{W}_2^{\mu}(s;\lambda)$ from part 6 of Lemma 1. Then, $\frac{\hat{\beta}}{T} = \frac{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}^k - \overline{Z}^k}{T^{3/2}}\right)\left(\frac{X_t - \overline{X}}{T^{1/2}}\right)}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{X_t - \overline{X}}{T^{1/2}}\right)^2} \Rightarrow \beta \frac{\int_0^{1-\lambda} \overline{W}_2^{\mu}(s;\lambda) W_2^{\mu}(s;\lambda) ds}{\int_0^{1-\lambda}\left(W_2^{\mu}(s;\lambda)\right)^2 ds}$. Moreover, $\frac{s^2}{T^3} = \frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1} - \overline{Z}^k}{T^{3/2}}\right)^2 - \left(\frac{\hat{\beta}}{T}\right)^2 \frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{X_t - \overline{X}}{T^{1/2}}\right)^2 \Rightarrow \beta^2\omega^2 \left[\frac{\int_0^{1-\lambda}\left(W_2^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds - \left(\int_0^{1-\lambda} W_2^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds\right)^2}{\int_0^{1-\lambda}\left(W_2^{\mu}(s;\lambda)\right)^2 ds}\right]$.

Under the null of $\beta = \beta_0$, the t-statistic is, $t = \frac{\left(\hat{\beta} - 0\right)\left(\sum_{t=1}^{T-k}\left(X_t - \overline{X}\right)^2\right)^{1/2}}{s}$.

We have to normalize it by $T^{1/2}$ to get: $\frac{t}{\sqrt{T}} = \frac{\left(\frac{\hat{\beta} - \beta}{T}\right)\left(\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{X_t - \overline{X}}{T^{1/2}}\right)^2\right)^{1/2}}{\left(\frac{s^2}{T^3}\right)^{1/2}} \Rightarrow$

$$\frac{\int_0^{1-\lambda} W_2^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds}{\left[\int_0^{1-\lambda}\left(W_2^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds - \left(\int_0^{1-\lambda} W_2^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds\right)^2\right]^{1/2}} = F_2\left(W_2^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right).$$

Finally, $R^2 = \left(\hat{\beta}\right)^2 \frac{\sum_{t=1t}^{T-k}\left(X_t - \overline{X}\right)^2}{\sum_{t=1}^{T-k}\left(Z_{t+1}^k - \overline{Z}^k\right)^2} =$

$$\left(\frac{\hat{\beta}}{T}\right)^2 \frac{\frac{1}{T}\sum_{t=1t}^{T-k}\left(\frac{X_t - \overline{X}}{T^{1/2}}\right)^2}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Z_{t+1}^k - \overline{Z}^k}{T^{3/2}}\right)^2} \Rightarrow \frac{\left(\int_0^{1-\lambda} W_2^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds\right)^2}{\int_0^{1-\lambda}\left(W_2^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds} = F_3\left(W_2^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right).$$

■

Proof of Theorem 5: This case deserves more attention. First, notice that $Z_{t+1}^k - \overline{Z}^k = \beta\left(Q_t^k - \overline{Q}^k\right) + R_t - \overline{R}$, where $R_t = \sum_{i=0}^{k-1} \varepsilon_{t+i+1}$ and $\overline{R} = \frac{1}{T}\sum_{t=0}^{T-k} R_t$. Also, $\frac{R_t - \overline{R}}{T^{1/2}} \Rightarrow W_1^{\mu}(s;\lambda)$. The OLS estimator is: $\hat{\beta} = \frac{\sum_{t=1}^{T-k}\left(Z_{t+1}^k - \overline{Z}^k\right)\left(Q_t^k - \overline{Q}^k\right)}{\sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}^k\right)^2} = \beta + \frac{\sum_{t=1}^{T-k}\left(R_t^k - \overline{R}\right)\left(Q_t^k - \overline{Q}^k\right)}{\sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}^k\right)^2}$. Therefore, $T\left(\hat{\beta} - \beta\right) = \frac{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{R_t^k - \overline{R}}{T^{1/2}}\right)\left(\frac{Q_t^k - \overline{Q}^k}{T^{3/2}}\right)}{\frac{1}{T}\sum_{t=1}^{T-k}\left(\frac{Q_t^k - \overline{Q}^k}{T^{3/2}}\right)^2} \Rightarrow \frac{\sigma_{11}}{\omega}\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds}{\int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds} = \frac{\sigma_{11}}{\omega} F_1\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right).$

Using similar steps, we can show that $\frac{t}{\sqrt{T}} \Rightarrow$

$$\frac{\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds}{\left[\int_0^{1-\lambda}\left(W_1^{\mu}(s;\lambda)\right)^2 ds \int_0^{1-\lambda}\left(\overline{W}_2^{\mu}(s;\lambda)\right)^2 ds - \left(\int_0^{1-\lambda} W_1^{\mu}(s;\lambda)\overline{W}_2^{\mu}(s;\lambda) ds\right)^2\right]^{1/2}} = F_2\left(W_1^{\mu}(s;\lambda), \overline{W}_2^{\mu}(s;\lambda)\right).$$

Lastly, $R^2 = \hat{\beta}^2 \frac{\sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}\right)^2}{\sum_{t=1}^{T-k}\left(Z_{t+1}^k - \overline{Z}^k\right)^2} = \frac{\hat{\beta}^2 \sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}\right)^2}{\beta^2 \sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}\right)^2 + 2\beta \sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}\right)\left(R_t^k - \overline{R}\right) + \sum_{t=1}^{T-k}\left(R_t^k - \overline{R}\right)^2}$. Dividing the numerator and the denominator by $\beta^2 \sum_{t=1}^{T-k}\left(Q_t^k - \overline{Q}\right)^2$, we obtain $R^2 = \frac{\frac{\hat{\beta}^2}{\beta^2}}{1 + o_p(1) + o_p(1)} \to^p 1$.■

References

[1] Barberis, N., 1999, Investing for the long run when returns are predictable, Journal of Finance, forthcoming.

[2] Brennan, M.J., E.S. Schwartz, and R. Lagnado, 1997, Strategic asset allocation, Journal of Economic Dynamics and Control 21, 1377-403.

[3] Campbell, J.Y., 1993, Why long horizons: A study of power against persistent alternatives, NBER Technical Working Paper No. 142.

[4] Campbell, J.Y. and J.H. Cochrane, 1999, By force of habit: A consumption based explanation of aggregate stock market behavior, Journal of Political Economy 107:2, 205-251.

[5] Campbell, J.Y. and L.M. Viceira, 1999, Consumption and poertfolio decisions when expecter returns are time varying, Quarterly Journal of Economics 114.

[6] Campbell, J.Y. and A. Lo and A.C. MacKinlay, 1997, The econometrics of financial markets, Princeton University Press.

[7] Campbell, J.Y. and R. Shiller, 1987, Cointegration and tests of present value models, Journal of Political Economy, 95:9, 1062-1088.

[8] Campbell, J.Y. and R. Shiller, 1988, The Dividend-Price ratio and expecations of future dividends and discount factors, Review of Financial Studies, 1:3, 195-228.

[9] Cavanagh, C., G. Elliott, and J. Stock, 1995, Inference in models with nearly integrated regressors, Econometric Theory, (11), 1131-1147.

[10] Cochrane, J.H., 1999, New facts in finance, manuscript, University of Chicago.

[11] Fama, E. and K. French, 1988, Dividend yields and expected stock returns, Journal of Financial Economics 22, 3-25.

[12] Fisher, M.E, and J.J. Seater, 1993, Long-Run Neutrality and Superneutrality in an ARIMA Framework, American Economic Review, 83:3, 402-415.

[13] Goetzmann, W. and P. Jorion, 1993, Testing the Predictive Power of Dividend Yields, Journal of Finance 48:2, 663-679.

[14] Granger, C.W.J. and P. Newbold, 1974, Spurious regressions in econometrics, Journal of Econometrics 2, 111-120.

[15] Hansen, L.P and R. Hodrick, 1980, Forward exchange rates as optimal predictors of future spot rates, Journal of Political Economy 88, 829-853.

[16] Hansen, B., 1992, Convergence to stochastic integrals for dependent heterogeneous processes, Econometric Theory, 8, 485-500.

[17] Hodrick, R., 1992, Dividend yields and expected stock returns: alternative procedures for inference and measurement, Review of Financial Studies (5:3), 357-386.

[18] Liu, J., 1999, Portfolio selection in stochastic environments, manuscript, Stanford University.

[19] Mishkin, F.S., 1990, What does the term structure of interest rates tell us about future inflation?, Journal of Monetary Economics 70, 1064-1072.

[20] Mishkin, F.S., 1992, Is the Fisher Effect for real?, Journal of Monetary Economics, 1992, 195-215.

[21] Nelson, C. and M. Kim, 1993, Predictable stock returns: the role of small sample bias, Journal of Finance (48:2), 641-661.

[22] Ng, S. and P. Perron, 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.

[23] Ng, S. and P. Perron, 1998, Lag length selection and the construction of unit root tests with good size and power, manuscript.

[24] Phillips, P.C.B., 1986, Understanding Spurious Regressions in Econometrics, Journal of Econometrics 33, 311-340.

[25] Phillips, P.C.B., 1987, Toward a unified asymptotic theory of autoregression, Biometrika, 74:3, 535-547.

[26] Phillips, P.C.B., 1991, Optimal Inference in Cointegrated Series, Econometrica, 59:2, 283-30.

[27] Richardson, Matthew and James H. Stock, 1989, Drawing inferences from statistics based on multi-year asset returns, Journal of Financial Economics 25, 323-348.

[28] Schwert, G.W., 1990, Indexes of the United States stock prices from 1802 to 1987, Journal of Business 63, 399-426.

[29] Siegel, J.J., 1992, The real rate of interest from 1800-1990:A study of the U.S. adn U.K., Journal of Monetary Economics 29, 227-52.

[30] Stambaugh, R., 1986, Bias in Regressons with Lagged Stochastic Regressors, Chicago working paper.

[31] Stambaugh, R., 1999, Predictive Regressons, Journal of Financial Economics, forthcoming.

[32] Stock, J.H., 1991, Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, Journal of Monetary Economics 28, 435-459.

[33] Stock, J., 1994, Unit roots, structural breaks and trends, in Engle, R. and D. McFadden (eds.), Handbook of Econometrics, Vol IV, Amsterdam: Elsevier, Chapter 46.

[34] Stock, J. and M. Watson, 1989, Interpreting the Evidence on Money-Income Causation, Journal of Econometrics 40: 1, 161-82.

[35] Stock, J. and M. Watson, 1993, A simple estimator of cointegrating vectors in higher order integrated systems, Econometrica 61:4, 783-820.

[36] Valkanov, R., 1998, The term structure with highly persistent interest rates, manuscript, Princeton University.

[37] Valkanov, R., 1999, Equity premium and dividend yield regressions: a lot of noise, little information, confusing results, manuscript, Princeton University.

[38] Watson, M., 1994, Vector Autoregressions and Cointegration, Handbook of Econometrics, Volume IV, R. Engle and D. McFadden eds, Elsevier Science, Chapter 47.

Table 3a

	Case 1			Case 2		
	T= 100	T= 750	Ratio	T= 100	T= 750	Ratio
$E(\hat{\beta})$	-0.0029	0.0069	-2.3566	-0.0006	0.0001	-0.2112
$Var(\hat{\beta})$	0.0868	0.0932	1.0740	0.0014	0.0000	0.0188
$E(t-stat.)$	-0.0478	0.1590	-3.3251	-0.0512	0.2096	-4.0968
$Var(t-stat.)$	10.0627	79.6769	7.9181	11.7233	91.0426	7.7660
$E(R^2)$	0.0862	0.0893	1.0365	0.0961	0.0989	1.0291
$Var(R^2)$	0.0107	0.0113	1.0519	0.0134	0.0134	1.0027

	Case 3			Case 4		
	T= 100	T= 750	Ratio	T= 100	T= 750	Ratio
$E(\hat{\beta})$	7.8246	56.8151	7.2611	0.9994	1.0001	1.0007
$Var(\hat{\beta})$	2.4431	159.9102	65.4540	0.0014	0.0000	0.0188
$E(t-stat.)$	16.8051	47.5789	2.8312	-0.0512	0.2096	-4.0968
$Var(t-stat.)$	80.7647	601.6150	7.4490	11.7233	91.0426	7.7660
$E(R^2)$	0.7264	0.6951	0.9569	0.9875	0.9998	1.0124
$Var(R^2)$	0.0304	0.0358	1.1792	0.0002	0.0000	0.0004

Notes: The system $(1-2)$ is simulated 5000 times, using samples of length $T = 100$ and $T = 750$. In cases 1 and 2, $\beta = 0$, whereas in cases 3 and 4, we let $\beta = 1$. The error terms are simulated from a standard normal distribution, and $\delta = 0$. Long-horizon series are produced as $Z_t^k = \sum_{i=0}^{k-1} Y_{t+i}$ and $Q_t^k = \sum_{i=0}^{k-1} X_{t+i}$, where $k = [\lambda T]$ and $\lambda = 0.1$. In each case and for each simulation, we estimate $\hat{\beta}$, its t-statistic under the appropriate null hypothesis, and the R^2. The first two moments of those three statistics are tabulated, for $T = 100$ and $T = 750$. The ratio of the moments is given in the last column of each table.

Table 3b

	Case 1			**Case 2**		
	T= 100	T= 750	Ratio	T= 100	T= 750	Ratio
$E(\hat{\beta})$	-0.1738	-0.2244	1.2906	0.0032	0.0000	0.0068
$Var(\hat{\beta})$	0.0855	0.0932	1.0901	0.0012	0.0000	0.0180
$E(t-stat.)$	-1.8886	-6.5913	3.4901	0.2731	0.0229	0.0838
$Var(t-stat.)$	10.0062	77.5463	7.7498	10.3546	74.4857	7.1935
$E(R^2)$	0.1083	0.1238	1.1438	0.0881	0.0842	0.9559
$Var(R^2)$	0.0155	0.0180	1.1594	0.0113	0.0104	0.9241

	Case 3			**Case 4**		
	T= 100	T= 750	Ratio	T= 100	T= 750	Ratio
$E(\hat{\beta})$	7.6463	56.2755	7.3599	1.0032	1.0000	0.9968
$Var(\hat{\beta})$	2.7653	168.3164	60.8668	0.0012	0.0000	0.0180
$E(t-stat.)$	15.1551	46.6912	3.0809	0.2731	0.0229	0.0838
$Var(t-stat.)$	65.9799	649.2143	9.8396	10.3546	74.4857	7.1935
$E(R^2)$	0.6904	0.6837	0.9902	0.9877	0.9998	1.0122
$Var(R^2)$	0.0358	0.0377	1.0525	0.0002	0.0000	0.0004

Notes: The system $(1-2)$ is simulated 5000 times, using samples of length $T = 100$ and $T = 750$. In cases 1 and 2, $\beta = 0$, whereas in cases 3 and 4, we let $\beta = 1$. The error terms are simulated from a standard normal distribution. Here, we let $\delta = 0.9$. Long-horizon series are produced as $Z_t^k = \sum_{i=0}^{k-1} Y_{t+i}$ and $Q_t^k = \sum_{i=0}^{k-1} X_{t+i}$, where $k = [\lambda T]$ and $\lambda = 0.1$. In each case and for each simulation, we estimate $\hat{\beta}$, its t-statistic under the appropriate null hypothesis, and the R^2. The first two moments of those three statistics are tabulated, for $T = 100$ and $T = 750$. The ratio of the moments is given in the last column of each table.

Table 4

Period 1927-1986

horizon	Dividend Yield: D(t)/P(t-1) $\hat\beta$	t	$t/\sqrt{T}$	percentile	Dividend Price ratio: D(t)/P(t) $\hat\beta$	t	$t/\sqrt{T}$	percentile
1 year	5.370	2.400	0.310	0.83	2.470	1.270	0.164	0.69
2 year	9.100	2.180	0.284	0.81	7.380	2.040	0.266	0.80
3 year	11.560	2.140	0.281	0.81	9.940	2.210	0.290	0.82
4 year	12.680	1.930	0.256	0.79	12.860	2.430	0.322	0.84

Period 1927-1956

horizon	Dividend Yield: D(t)/P(t-1) $\hat\beta$	t	$t/\sqrt{T}$	percentile	Dividend Price ratio: D(t)/P(t) $\hat\beta$	t	$t/\sqrt{T}$	percentile
1 year	11.040	2.490	0.455	0.92	1.500	0.460	0.084	0.60
2 year	22.490	2.880	0.535	0.95	8.920	1.490	0.277	0.81
3 year	29.240	2.860	0.540	0.95	15.270	2.210	0.418	0.90
4 year	28.160	2.250	0.433	0.91	20.860	3.140	0.604	0.97

Period 1957-1986

horizon	Dividend Yield: D(t)/P(t-1) $\hat\beta$	t	$t/\sqrt{T}$	percentile	Dividend Price ratio: D(t)/P(t) $\hat\beta$	t	$t/\sqrt{T}$	percentile
1 year	5.600	1.860	0.340	0.85	9.320	3.020	0.551	0.95
2 year	7.510	1.890	0.351	0.86	16.400	4.040	0.750	0.99
3 year	10.410	3.010	0.569	0.96	17.120	4.120	0.779	0.99
4 year	15.050	3.370	0.649	0.97	19.690	3.870	0.745	0.99

Period 1941-1986

horizon	Dividend Yield: D(t)/P(t-1) $\hat\beta$	t	$t/\sqrt{T}$	percentile	Dividend Price ratio: D(t)/P(t) $\hat\beta$	t	$t/\sqrt{T}$	percentile
1 year	4.460	2.620	0.386	0.89	5.090	2.880	0.425	0.91
2 year	7.150	3.040	0.453	0.92	10.340	4.180	0.623	0.97
3 year	9.420	4.770	0.719	0.98	12.940	5.680	0.856	0.99
4 year	12.750	5.490	0.837	0.99	15.350	5.620	0.857	0.99

Notes: The columns named $\hat\beta$ report the OLS estimate of regressing the long-horizon returns on the dividend yield or the dividend-price ratio. All the values in the table are taken or computed from Fama and French (1988), table 3. The third column is the appropriately normalized t-statistic. The last column reports the percentile of the normalized t-statistic, under the null of no relation between expected returns and dividend/price ratios, or $\beta = 0$ and $\delta = 0.04$, as discussed in the text . This long-horizon regression corresponds to case 1. Unlike the non-normalized t-statistic, which sujests a clear rejection of the null, the $t/\sqrt{T}$ statistic rejects the null at the 0.05-level, but cannot reject the null at the 0.01 (two sided test). Although the evidence of a long-horizon forecasting relationship is not as strong as suggested by Fama and French (1988), it is present, especially in the post WWII periods.

Table 5

Period 1802-1990

horizonl	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
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

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
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

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
5-years	0.432	-11.360	-1.303	0.01
IV:lagged 5 years	2.120	0.941	0.069	0.58

Notes: The first column reports the OLS estimate of regressing the 5-month nominal stock return on the 5-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" is taken from Boudoukh and Richardson (1993), table 2, case ii. The third column is the appropriately normalized t-statistic. The last column reports the percentile of the normalized t-statistic, under the null of a full Fisher effect (Mishkin (1992) and Gali (1988)), or $\beta = 1$. This long-horizon regression corresponds to case 4. The normalized t-statistic cannot reject the null in two out of the three periods for the OLS case, contrary to the conclusions reached from the un-normalized t-statistic. The null cannot be rejected in any period, for the IV case .

Table 6

Period 1953-1990

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
3 months	0.663	-5.507	-0.261	0.21

Period 1953-1979

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
3 months	1.188	2.642	0.150	0.66

Period 1979-1982

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
3 months	0.235	-3.027	-0.505	0.07

Period 1982-1990

horizon	$\hat{\beta}$	t	$t/\sqrt{T}$	percentile
3 months	0.125	-5.573	-0.569	0.05

Notes: The first column reports the OLS estimate of regressing the 3-month inflation on the 3-month interest rate. The values are taken from Mishkin (1992), table 1. The third column is the appropriately normalized t-statistic. The last column reports the percentile of the normalized t-statistic, under the null of a full Fisher effect (Mishkin (1992) and Gali (1988)), or $\beta = 1$. This long horizon regression corresponds to case 4. Unlike the t-statistic, the normalized t-statistic cannot reject the null. However, the evidence for a full Fisher effect is less convincing for the last two sub-periods. A similar conclusion was reached by Mishkin (tables 3-4), using Monte Carlo simulations.

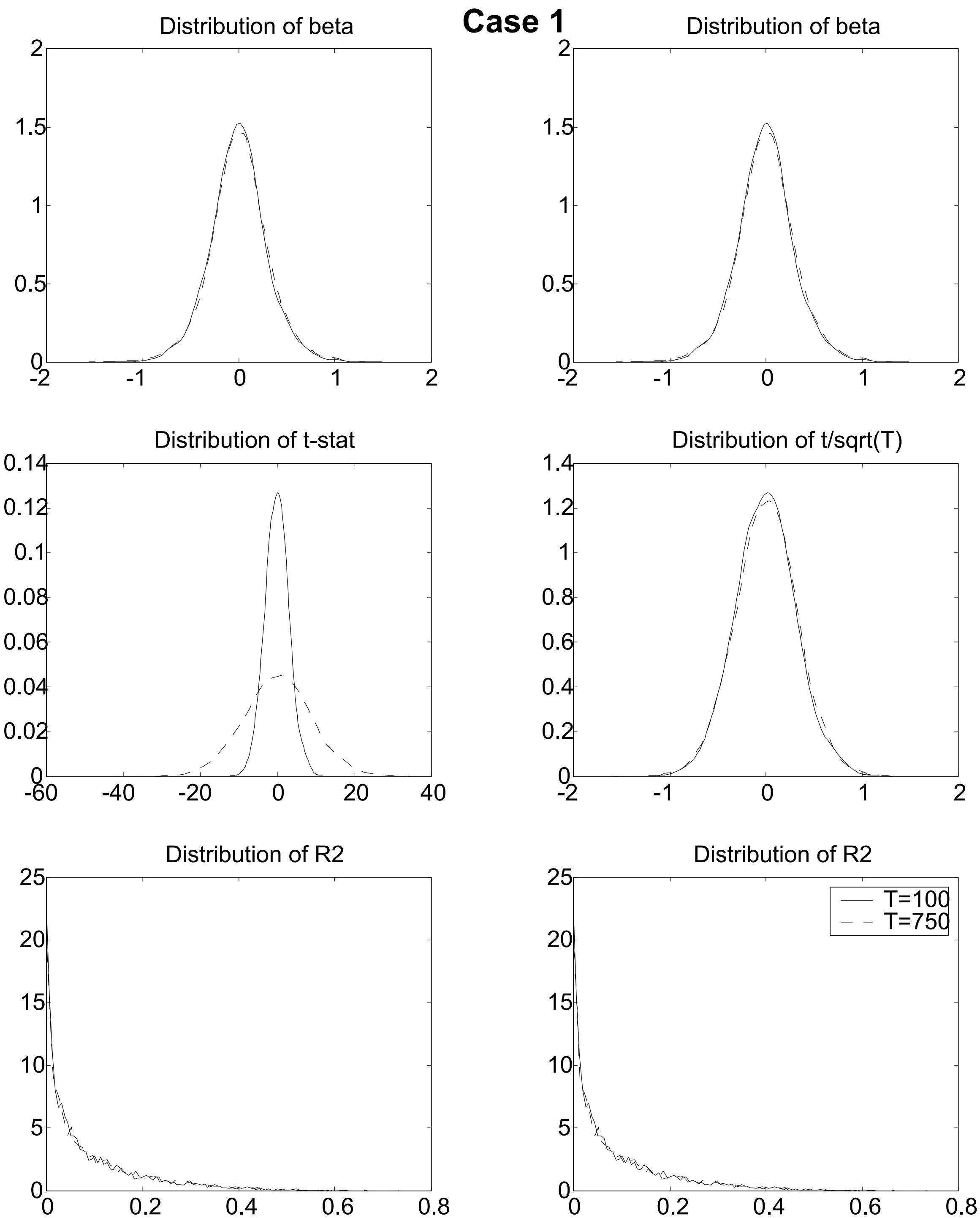

Figure 1: The graphs in the first column are the distributions of $\hat{\beta}$, the t-statistic under the null $\beta = 0$, and the R^2 in Case 1, $\delta = 0$, for T=100, and 750. The distributions of the same statistics, appropriately rescaled to converge asymptotically, are presented in the second column. The variance of $\hat{\beta}$ does not decrease as T increases, because the estimator is not consistent, as proven in Theorem 2. The variance of the t-statistic increases with the sample size, but $t/\sqrt{T}$ has a well defined asymptotic distribution, as expected from Theorem 2. The R^2 does not converge to 1, as T increases. Its distribution has a considerable mass at zero, but values of 0.4 are not unusual.

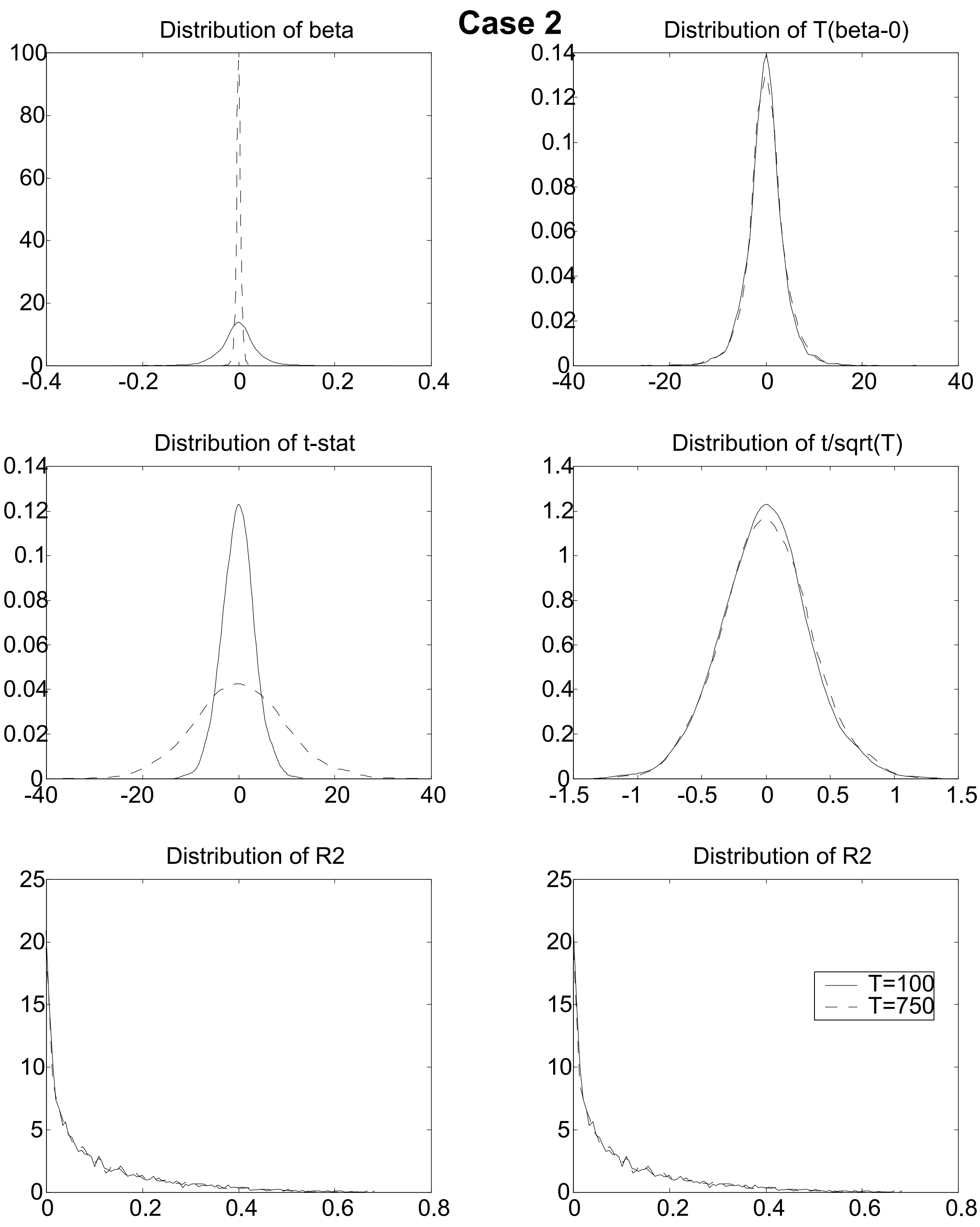

Figure 2: The graphs in the first column are the distributions of $\hat{\beta}$, the t-statistic under the null $\beta = 0$, and the R^2 in Case 2, $\delta = 0$, for T=100, and 750. The distributions of the same statistics, appropriately rescaled to converge asymptotically, are presented in the second column. The variance of $\hat{\beta}$ decreases at rate T^2, because the estimator is super-consistent, as proven in Theorem 3. The variance of the t-statistic increases with the sample size, but $t/\sqrt{T}$ has a well defined asymptotic distribution, as expected from Theorem 3. The R^2 does not converge to 1, as T increases. Its distribution has a considerable mass at zero, but values of 0.4 are not unusual.

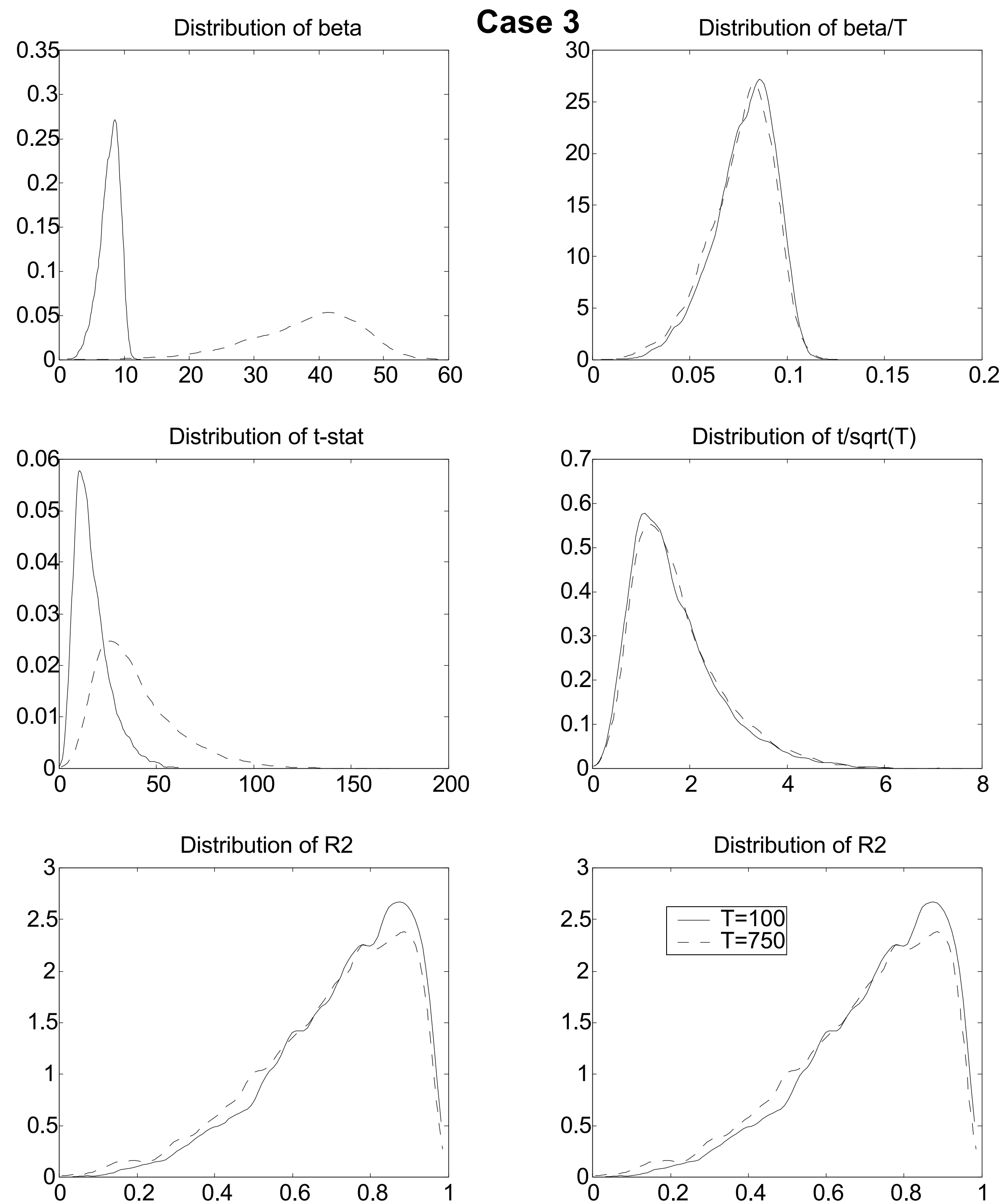

Figure 3: The graphs in the first column are the distributions of $\hat{\beta}$, the t-statistic under the null $\beta = 1$, and the R^2 in Case 3, $\delta = 0$, for T=100, and 750. The distributions of the same statistics, appropriately rescaled to converge asymptotically, are presented in the second column. The variance of $\hat{\beta}$ does not decrease as T increases, because the estimator is not consistent, as proven in Theorem 4. In fact, higher values of T or k (since k=$[\lambda T]$) result in higher estimates of $\hat{\beta}$. The variance of the t-statistic increases with the sample size, but $t/\sqrt{T}$ has a well defined asymptotic distribution, as expected from Theorem 4. The R^2 does not converge to 1, as T increases. Its distribution has a considerable mass away from zero.

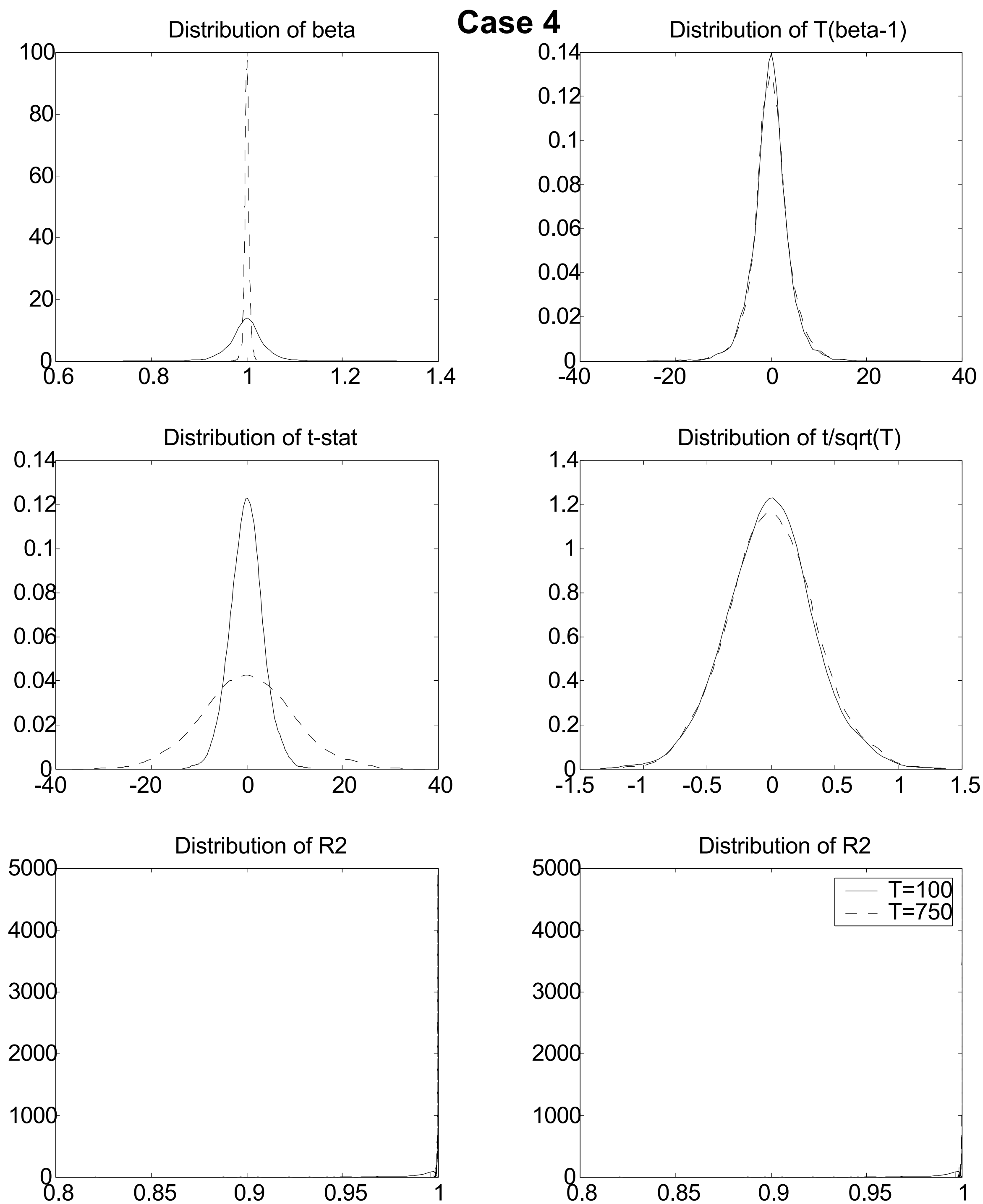

Figure 4: The graphs in the first column are the distributions of $\hat{\beta}$, the t-statistic under the null $\beta = 1$, and the R^2 in Case 4, $\delta = 0$, for T=100, and 750. The distributions of the same statistics, appropriately rescaled to converge asymptotically, are presented in the second column. The variance of $\hat{\beta}$ decreases at rate T^2, because the estimator is super-consistent, as proven in Theorem 5. The variance of the t-statistic increases with the sample size, but $t/\sqrt{T}$ has a well defined asymptotic distribution, as expected from Theorem 5. The R^2 does not converge to 1, as T increases. Its distribution has a considerable mass away from zero.

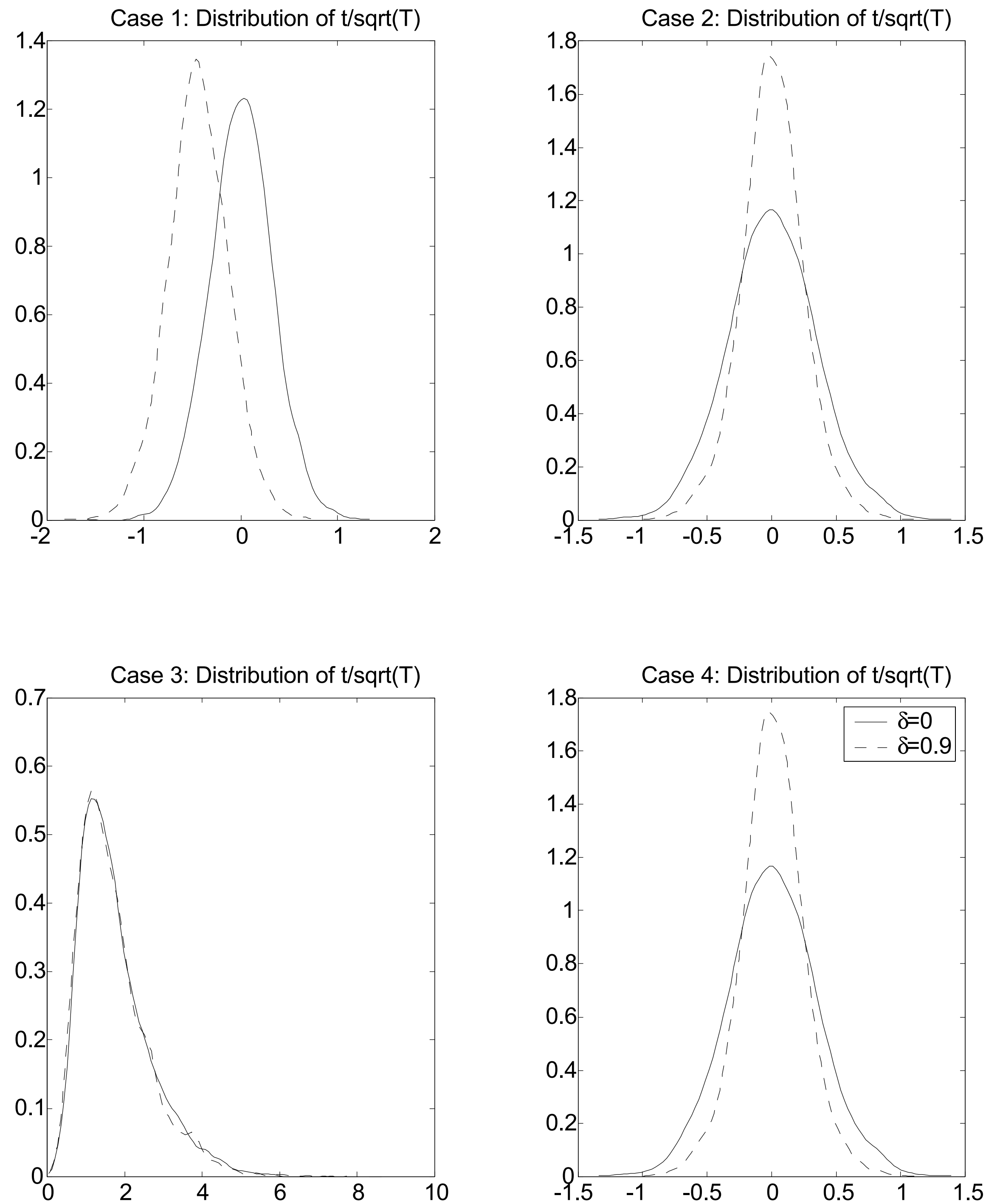

Figure 5: The asymptotic distributions of the rescaled statistics in all four cases are compared for $\delta = 0$ and $\delta = 0.9$. A correlation between $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ affects the mean and the variance of the distribution in the first case. The second and the fourth cases have identical asymptotic distributions. The variance only is affected. In case 3, the distribution of $t/\sqrt{T}$ is invariant to δ.