Mathematical Finance, Vol. 4, No. 2 (April 1994), 103–119

THE STATISTICS OF LONG-HORIZON REGRESSIONS REVISITED[1]

JACOB BOUDOUKH

Stern School of Business, New York University, New York, NY

MATTHEW RICHARDSON

Wharton School, University of Pennsylvania, Philadelphia, PA

This paper compares commonly used approaches for estimating the relation between long-horizon returns and a predetermined variable X_t, such as dividend yields. Specifically, we look at regressions of (i) nonoverlapping multiperiod returns on X_t, (ii) overlapping multiperiod returns on X_t, (iii) single-period returns on multiperiod X_t, and (iv) single-period returns on X_t and its implied long-horizon regression coefficient. We provide analytical formulae which quantify the efficiency of the estimators used in the various approaches. Using the formulae, as well as Monte Carlo simulations, we demonstrate that the relative efficiency of the estimators used in the various approaches differs remarkably, depending on the dynamic structure of the regressor. Of special interest for financial economists, when the regressors are highly autocorrelated, we find that the regressions (ii), (iii), and (iv) provide only *marginal* efficiency gains above and beyond the nonoverlapping long-horizon regression.

KEY WORDS: long-horizon regressions, time aggregation, asymptotic standard errors, relative efficiency

1. INTRODUCTION

A plethora of papers exist which investigate the relation between long-horizon stock returns and predetermined variables, such as current dividend yields, interest rate spreads, and volatility.[2] However, despite this volume of research, little is known about the motivation for long-horizon tests and, even more important, the statistical properties of these tests.[3] This paper fills a void in the existing literature by analyzing the statistics of long-horizon regressions. In particular, we focus our attention on several procedures for estimating long-horizon regressions currently used in the finance literature.

We investigate the claim that the use of overlapping observations in calculating long-horizon relations is far more efficient than using nonoverlapping observations. Clearly, at some level, this point should come as no surprise to econometricians since nonoverlapping data ignores information in the time series. Based on an asymptotic argument, Hansen and Hodrick (1980) provide support for this claim, in a fairly general setting. They provide, however, no discussion of the magnitude of the benefit from the use of overlapping observations. Richardson and Smith (1991) do manage to quantify the benefit when the regressor

[1] We would like to thank Bob Cumby, George Tauchen, Stephen Taylor, Rob Whitelaw, seminar participants at Duke University, and two anonymous referees for helpful comments and suggestions.

[2] See, for example, Kandel and Stambaugh (1988), Fama and French (1988, 1989), Cochrane (1991a), Bekaert and Hodrick (1992), Boudoukh and Richardson (1993), and Hodrick (1992), among others.

[3] Some recent exceptions are Campbell (1993b) and Stambaugh (1993), who describe advantages of long-horizon regressions in terms of their potential power relative to short-horizon regressions; and Goetzmann and Jorion (1993) and Nelson and Kim (1993), who describe some of the biases that arise in using long-horizon regressions. The relation of this work to those papers will be discussed in Section 4.

Manuscript received March 1993; final revision received September 1993.

© 1994 Blackwell Publishers, 238 Main Street, Cambridge, MA 02142, USA, and 108 Cowley Road, Oxford OX4 1JF, UK.

is a lagged value of the dependent variable. No results have been given, however, for the more current and more relevant literature regarding regressions of long-horizon returns on predetermined (i.e., instrumental) variables.

In a statistical setting of particular interest to financial economists, we analytically quantify the efficiency of using overlapping data and provide support for our results via simulation evidence. As a preview, our main result is that this efficiency is directly tied to the dynamic structure of the regressor. For example, with highly autocorrelated regressors (as is usually the case in monthly financial time series), overlapping observations buys the econometrician very little in the form of efficiency. For example, consider regressing five-year stock returns on its dividend yield using monthly data over a 60-year period. Suppose further that the autocorrelation of the predetermined variable is .98, as is the case for the monthly price-dividend ratio series. The benefit of using overlapping observations can be quantified as the ratio of the variance of the regression coefficient for the nonoverlapping versus the overlapping case. This ratio is, for our example, 1.44. Put differently, the use of approximately 60 times more observations buys us variances only 30.6% tighter. Similar results are obtained for finite samples, as is demonstrated via simulations. This suggests that researchers should be cautious in their interpretation of results in long-horizon settings and not be deceived by the apparently large number of observations available through the use of overlapping data.

In contrast, the gains to using overlapping data can be substantial when the autocorrelation of the regressor is reduced. For example, consider regressing five-year stock returns on one-year inflation rates using annual data over the past 200 years (see Boudoukh and Richardson 1993). Since the autocorrelation of inflation is approximately .5 over this period, it is possible to show that the ratio of variances of the estimators from using nonoverlapping versus overlapping data is 2.2. Roughly speaking, using 200 overlapping observations here is equivalent to using 88 nonoverlapping observations in terms of the efficiency of the estimator. While the gains to using overlapping data can clearly be large, researchers recognize its limitations. In particular, in terms of providing usual tests of significance, the drawback of performing regressions of long-horizon returns on predetermined variables is that it requires the estimation of the variance-covariance matrix of serially correlated errors (see Richardson and Smith 1991 and Hodrick 1992). Since this necessitates estimation of numerous autocovariances, the small-sample properties of these statistics suffers.

Therefore, some researchers have argued against this strategy in favor of an alternative method, which calls for the regression of single-horizon returns on the same predetermined variable, except that this predetermined variable is now calculated over long horizons. Hodrick (1992), for example, argues that this procedure can have better small sample properties while capturing the important information contained in the long-horizon method.[4] The idea is simple—this alternative method avoids the practical limitations of having to calculate standard errors in the presence of overlapping observations, since it falls within the standard ordinary least-squares framework. Furthermore, since both methods estimate the same multiperiod covariance between returns and the predetermined variable, there is apparently little cost in switching to the second method.

This claim can be misguided. It turns out that, while the calculation of standard errors are less subject to sampling error, there are potentially large costs to using the alternative

[4] Cochrane (1991a) also argues for transforming the long-horizon regression in this way.

method in terms of the small sample efficiency of the estimator itself. The intuitive explanation is that the alternative method involves calculating the multiperiod variance of the regressor while the original long-horizon method requires only the single-period variance. In small samples, this distinction can have profound effects. Consider, for example, comparing the original method in which we regress 10-year stock returns on a predetermined variable with an autocorrelation of .99, to the alternative method in which we regress monthly stock returns on 10-year aggregates of the predetermined variable, using data over a 60-year period. We show that the latter method provides variances 318.5% larger than the original method. It is fairly clear therefore that using the alternative procedure can lead to losses of efficiency in small samples. Thus, researchers should be especially cautious in applying this method to data.

We also provide some new discussion on the benefits and drawbacks of using a long-horizon regression versus a short-horizon regression approach. Recent research by Kandel and Stambaugh (1988) and Hodrick (1992) suggests that calculating implied long-horizon coefficients from short-horizon regressions of returns on predetermined variables may be preferred to estimating analogous coefficients from regressions of long-horizon returns on the same predetermined variables.

Our extension of this literature is to quantify the benefit of using the short-horizon approach. Our main result is that the short-horizon method can often provide remarkable gains in terms of efficiency, although these are only marginally beneficial for highly autocorrelated regressors. As an illustration, to coincide with previous examples, consider regressing five-year stock returns on a highly autocorrelated predetermined variable. The benefit to using the implied long-horizon coefficient relative to using the actual long-horizon coefficient can be essentially zero (in terms of the tightness of the variances). In contrast, consider the same regression using monthly data and assume that the predetermined variable is uncorrelated through time. The variance of the short-horizon approach is 1/60 that of overlapping multiperiod regressions, and 1/3600 that of nonoverlapping multiperiod regressions!

It is important to point out that the focus of this paper is the efficiency of long-horizon estimators under the null hypothesis of no predictability. The aforementioned results are theoretical, with support from empirical small sample distributions. Using the data to calculate a measure of efficiency, such as OLS standard error estimates, is besides the point. Either the standard error calculations reach the same conclusions as we have here or they are not good measures in small samples. For the case of highly autocorrelated regressors, researchers should therefore not be fooled by the apparently large number of observations, and should be wary of small estimated standard errors (the magnitude of the variances of the regressor and regressand aside).

This paper is organized as follows. Section 2 provides the statistical setting for our analysis. Section 3 gives the main results of the paper. Specifically, we provide analytical results and supporting simulation evidence with respect to each of the aforementioned claims in the literature. In Section 4, we discuss some related work in this area and provide directions for future research.

2. STATISTICAL FRAMEWORK

In this section, we lay out the statistical setting of the long-horizon regression problem looked at throughout the paper. There has been substantial interest in the finance literature at trying to understand the relation between long-horizon returns and predetermined variables. Let us denote the coefficient describing this relation in a J- period regression context as β_J. Some examples of this research (although not exhaustive of the literature) include Campbell and Shiller (1987), Fama and Bliss (1987), Kandel and Stambaugh (1988), Fama and French (1988, 1989), Cochrane (1991a,b), Bekaert and Hodrick (1992), Hodrick (1992), Boudoukh and Richardson (1993), Campbell (1993a), Goetzmann and Jorion (1993), and Nelson and Kim (1993). In terms of the predictability of returns, authors have found more significant economic and statistical evidence at long horizons (e.g., see Fama and Bliss 1987; Fama and French 1988; Hodrick 1992).[5] Across this large volume of research, a variety of procedures has been developed for inferring long-horizon relations. Although the statistical procedures vary across these papers, the initial focus had been on understanding the regression

$$R_{t,t+J} = \alpha_J^{\text{nol}} + \beta_J^{\text{nol}} X_{t,J} + \epsilon_{t,J}^{\text{nol}}, \tag{2.1}$$

where $R_{t,t+J}$ = continuously compounded return from t to $t + J$, sampled every J periods,
$X_{t,J}$ = predetermined variable, sampled every J period,
β_J^{nol} = regression coefficient using nonoverlapping observations.

Equation (2.1) describes a regression of long-horizon returns on a predetermined variable, X_t, using nonoverlapping data. Prior to the work of Hansen and Hodrick (1980), regressions like (2.1) were the standard approach to estimating multiperiod systems, due to the simple dynamic structure of the error term, $\epsilon_{t,J}^{\text{nol}}$. However, during the past decade, regression equation (2.1) has been modified in an attempt to improve the efficiency of estimators of β_J. This paper provides a quantitative analysis of the relative efficiency of these modified procedures.

Before proceeding with our analysis, it is necessary to make some additional simplifying assumptions:

- We look at the properties of the regression coefficients under the null hypothesis of no predictability, i.e., $\beta_J = 0$. We do this to coincide with the majority of empirical studies which focus on this hypothesis. Extensions to alternative hypotheses, i.e., $\beta_J \neq 0$, are straightforward, though cumbersome, in our framework (see Campbell 1993b and Stambaugh 1993).
- We investigate (2.1) under the assumption that the predetermined variable follows a first-order autoregressive process

$$X_t = \mu + \rho X_{t-1} + \eta_t, \tag{2.2}$$

where η_t is an i.i.d. mean-zero random variable. Campbell (1993b), Stambaugh (1993), and others have found this specification particularly useful in illustrating the predictability of stock returns. Moreover, for many of the relevant financial data series, an

[5] These conclusions, however, are not entirely uncontroversial; for example, Goetzmann and Jorion (1992) and Nelson and Kim (1991) question the small-sample properties of long-horizon regressions. In an interesting paper, Foster and Smith (1992) suggest that much of the evidence on predictability is overstated, basing their argument on a formal analysis of data snooping.

AR(1) seems to be fairly representative. Extensions to more elaborate ARMA processes are possible in this setting, and are left for future work.

- For convenience, we focus only on univariate X_t. Extensions to multivariate X_t, which could for example include dividend yields, volatility, and interest rates in the same regression, are potentially interesting. However, in terms of understanding the statistics of long-horizon regressions as implied by the various procedures, the extension is of less importance.
- We do not allow for heteroskedasticity in the disturbance terms in (2.1) or (2.2). Heteroskedasticity is not the focus of our paper and is thus beyond its immediate scope. A detailed analysis is provided by Stambaugh (1993), and his work is discussed in Section 4.

Throughout the paper, we explore the statistical properties of regressions analogous to (2.1). Our general approach will be to derive asymptotic analytical comparisons of the various methodologies. Specifically, we compare the asymptotic variances of the estimators from the various methods under the null hypothesis of no predictability.[6] To provide some supporting evidence of small-sample behavior of the regression coefficients, we employ a Monte Carlo simulation method. In particular, for each combination of J and ρ, we conduct 1000 replications of each simulation under the null hypothesis that the predetermined variable is independent of the dependent variable. The errors are generated from a mean-zero normal distribution, where the unconditional variance of the predetermined variable is kept fixed.

3. ECONOMETRICS OF LONG-HORIZON REGRESSIONS

3.1. Overlapping versus Nonoverlapping Observations

Consider the following alternative regression to (2.1):

$$\sum_{i=1}^{J} R_{t+i} = \alpha_J^{\text{ol}} + \beta_J^{\text{ol}} X_t + \epsilon_t(J), \tag{3.1}$$

where
R_t = continuously compounded return from $t-1$ to t,
X_t = predetermined variable, such as a portfolio's dividend yield,
β_J^{ol} = regression coefficient using overlapping observations,
$\epsilon_t(J)$ = J-sum of mean zero uncorrelated disturbance terms: $\epsilon_t + \epsilon_{t+1} + \cdots + \epsilon_{t+J-1}$, under the null hypothesis $\beta_J^{\text{ol}} = 0$.

The only difference between (2.1) and (3.1) is that the latter uses overlapping data while the former simply samples the data every J periods. That is, (3.1) employs T observations

[6] This comparison has asymptotic justification via the approximate slope procedure of Bahadur (1960) and Geweke (1981). In particular, results in those papers imply that, for this particular example, the ratio of variances roughly corresponds to the ratio of observations needed to achieve the same power under *any* alternative. Hence, if the ratio, $\text{var}(\hat{\beta}_J^{\text{nol}})/\text{var}(\hat{\beta}_J^{\text{alt}})$, is 2 to 1, then approximately twice as many observations are required from the nonoverlapping approach, where $\hat{\beta}_J^{\text{alt}}$ is the coefficient estimator from the alternative method. For more extensive discussion of the approximate slope procedure as applied in finance, see Richardson and Smith (1991, 1992) and Campbell (1993b).

in estimation, while (2.1) uses only T/J observations. Intuition suggests that (3.1) will lead to more efficient estimates of the coefficient describing the regression of J-period returns on X_t. This intuition is confirmed quite generally in Hansen and Hodrick (1980), who show that overlapping data will lead to asymptotically more efficient estimators. They justify their results by appealing to the approximate slope procedure of Bahadur (1960) and Geweke (1981). In addition, their results have been quantified by Richardson and Smith (1991), who show that, in the special case of estimating multiperiod autocorrelations, overlapping observations provide approximately 50% more observations. As yet, no quantifiable comparison has been performed for the more general multivariate setting described by (2.1) and (3.1). However, this approach of Richardson and Smith (1991) is straightforward to apply in this setting.

To see this, note that the normal equations from regression (3.1) are simply

$$E[f_t(\cdot)] \equiv E\begin{pmatrix} \epsilon_t(J) \\ \epsilon_t(J)X_t \end{pmatrix} = 0.$$

Under the null that $\beta_J = 0$, and applying results from Hansen (1982), the vector of regression coefficients $\hat{\theta} = (\hat{\alpha}_J^{\mathrm{ol}}\ \hat{\beta}_J^{\mathrm{ol}})'$ has an asymptotic normal distribution with mean $(\alpha_J \quad 0)'$ and variance-covariance matrix $[D_0' S_0^{-1} D_0]^{-1}$, where $D_0 = E[\partial f_t/\partial\theta]$ and $S_0 = \Sigma_{i=-\infty}^{+\infty} E[f_t f_{t-i}']$. Under the assumptions described in Section 2, and for a given overlap of J, it is possible to calculate D_0 and S_0 analytically. Specifically,

$$D_0 = -\begin{pmatrix} 1 & \mu_x \\ \mu_x & \mu_x^2 + \sigma_x^2 \end{pmatrix},$$

$$S_0 = \begin{pmatrix} J^2\sigma_R^2 & J^2\sigma_R^2\mu_x \\ J^2\sigma_R^2\mu_x & J^2\sigma_R^2\mu_x + \sigma_R^2\sigma_x^2\left[J + \dfrac{2\rho}{1-\rho}\left(J - 1 - \rho\dfrac{1-\rho^{J-1}}{1-\rho}\right)\right] \end{pmatrix},$$

where μ_x is the mean of X_t, σ_x^2 is the unconditional variance of X_t, σ_R^2 is the variance of single period returns R_t, and recall that ρ is the AR(1) parameter on X_t.

Using D_0 and S_0 above, and performing the necessary matrix calculations, it is possible to derive the asymptotic variance of $\hat{\beta}_J^{\mathrm{ol}}$:

$$\text{(3.2)} \qquad \operatorname{var}(\hat{\beta}_J^{\mathrm{ol}}) = \frac{\sigma_R^2}{\sigma_x^2}\left[J + \frac{2\rho}{1-\rho}\left(J - 1 - \rho\frac{1-\rho^{J-1}}{1-\rho}\right)\right].$$

Note that as the length of the horizon J increases, so does the variance of the regression coefficient, $\hat{\beta}_J^{\mathrm{ol}}$. More interesting is that, as ρ increases (i.e., as X_t approaches a random walk), the variance also increases. This suggests that for highly autocorrelated regressors (as is usually the case in financial series) the precision of the estimator in the presence of overlapping data vastly diminishes.

In contrast, with respect to regression equation (2.1), and using straight OLS theory, it is possible to calculate the asymptotic variance of the regression coefficient, $\hat{\beta}_J^{\mathrm{nol}}$. Noting that only T/J observations are now used in estimation and thus making the necessary adjustment for comparison with (3.2), we find that

$$\text{var}(\hat{\beta}_J^{\text{nol}}) = J^2 \sigma_R^2 / \sigma_x^2. \tag{3.3}$$

Comparing the variances of $\hat{\beta}_J^{\text{nol}}$ and $\hat{\beta}_J^{\text{ol}}$ leads to the immediate result that using overlapping observations is more efficient than using nonoverlapping observations. Table 3.1A provides the relative efficiency of the overlapping versus nonoverlapping method for different horizon length J and different autocorrelation parameter ρ. For example, consider the case in which X_t is independent through time; i.e., $\rho = 0$. As one might expect, the advantage of using overlapping observations is large in this instance. At $J = 60$, overlapping regression equation (3.1) is 6000% more efficient than nonoverlapping regression equation (2.1). This result indicates that using overlapping data provides the econometrician with an enormous efficiency gain in this instance, and more generally with predetermined variables characterized by low persistence. Put differently, for $\rho = 0$, the J-period overlapping regression provides roughly J times more observations than the nonoverlapping regression.

In contrast, this benefit is almost completely eliminated for highly autocorrelated X_t. For example, for $J = 60$ and $\rho = .99$, overlapping regression equation (3.1) is only 21% more efficient than nonoverlapping regression equation (2.1). With respect to applications of long-horizon regressions in finance, this is an especially interesting result for several reasons. First, most predetermined variables (which are used in long-horizon return regressions) are highly autocorrelated. For example, for monthly data, popular instrumental variables, such as dividend yields, corporate bond spreads, and interest rate levels, have first-order autocorrelations of .98, .97, and .99, respectively (see, for example, Whitelaw 1992). Thus, as seen in Table 3.1A for highly autocorrelated regressors, the use of overlapping observations provides only marginal efficiency benefits.

Second, there are costs to using overlapping observations in practice. Due to the induced serial correlation in the OLS errors, estimation procedures using overlapping observations require calculation of a consistent estimator of the variance-covariance matrix. This calculation involves numerous estimates of autocovariances at various lags, which can lead to difficulties in small samples for test statistic applications (see Richardson and Smith 1991).[7] Therefore, given the highly autocorrelated regressors used in practice, and the difficulties with overlapping observations, researchers should be cautious in their application of long-horizon regressions.

For example, consider using 720 monthly observations to forecast five-year returns. If the forecasting variable, X_t, were uncorrelated through time, then the overlapping approach would be as efficient as a nonoverlapping regression which uses 720 five-year independent observations, or roughly 3600 years of data! However, if the autocorrelation of X_t were .99, then the overlapping approach would be as efficient as a nonoverlapping regression which uses 14 five-year independent observations, or only 10 more years of data than we had initially. Since most researchers would be a priori cautious of regressions with only

[7] It is possible to avoid these difficulties in some circumstances by calculating the variance-covariance matrix directly under the null as in (3.2). For more detailed discussions of this approach, see Richardson and Smith (1991) and extensions to multivariate data in Hodrick (1992). However, as pointed out by Campbell (1993), for some alternative models of interest, it may be beneficial not to impose the null $\beta_J = 0$ when estimating standard errors.

TABLE 3.1
Nonoverlapping, Overlapping, and Implied Regressions–Asymptotic Results

J	$\rho = 0.00$	$\rho = 0.50$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
		A: Nonoverlapping versus Overlapping			
1	1.00	1.00	1.00	1.00	1.00
2	2.00	1.33	1.05	1.03	1.01
12	12.00	4.50	1.46	1.21	1.04
24	24.00	8.47	1.98	1.45	1.08
60	60.00	20.45	3.75	2.23	1.21
120	120.00	40.45	6.86	3.67	1.44
		B: Nonoverlapping versus Implied			
1	1.00	1.00	1.00	1.00	1.00
2	4.00	1.77	1.10	1.06	1.02
12	144.00	36.00	3.38	1.69	1.11
24	576.00	143.99	6.79	2.87	1.25
60	3600.00	899.80	36.15	9.90	1.75
120	14400.00	3600.05	144.06	36.15	2.94

The table provides a comparison of the relative efficiency of the long-horizon regression coefficient using three methods: (i) overlapping observations, (ii) nonoverlapping observations, and (iii) the implied coefficient from short-horizon regressions under AR(1). The estimation involves the regression of a J-period return on a predetermined variable, X_t, which follows a first-order autoregressive process (with corresponding parameter ρ). Panel A provides the analytical asymptotic ratio between the variance of the nonoverlapping regression coefficient and the overlapping regression coefficient for various values of J and ρ. Panel B provides the analytical asymptotic ratio between the variance of the nonoverlapping regression coefficient and the implied regression coefficient.

14 observations, the evidence here puts into question long-horizon regressions as a viable alternative approach.

The analytical results in Table 3.1A suggest that the strategy of using overlapping observations is highly sensitive to the degree of autocorrelation of the predetermined variable. These analytical calculations, however, are only strictly valid asymptotically. It seems worthwhile exploring whether the intuition carries through to small samples. Using the Monte Carlo simulation described in Section 2, we compare the small sample variances of $\hat{\beta}_J^{ol}$ and $\hat{\beta}_J^{nol}$, as calculated from their respective Monte Carlo empirical distributions. In particular, to coincide with existing studies, we choose the number of observations T equal to 720; the level of the autocorrelation parameter ρ to take the values (0, .5, .9, .95, .99); and the horizon length J to take the values (2, 12, 24, 60, 120). The results for these simulations are provided in Table 3.2A.

The results in Table 3.2A generally support the analytical calculations given in Table 3.1A. In particular, there are tremendous gains from using overlapping observations for independent X_t, but these gains greatly diminish as we increase the autocorrelation parameter. For example, for $\rho = 0$, the ratio of variances is 2.02, 13.99, 28.41, 70.06, and 260.18 when J equals 2, 12, 24, 60, and 120 respectively. In contrast, for $\rho = .99$, the ratio is only 1.02, 1.04, 1.14, 1.30, and 2.72 when J equals 2, 12, 24, 60, and 120 respectively.

Note that, for J equal to 60 and 120, the benefits are slightly higher in Table 3.2A than those implied by the analytical results in Table 3.1A. However, this is not due to increased

TABLE 3.2
Nonoverlapping, Overlapping, and Implied Regressions—Simulation Results

J	$\rho = 0.00$	$\rho = 0.50$	$\rho = 0.9$	$\rho = 0.95$	$\rho = 0.99$
		A: Nonoverlapping versus Overlapping			
1	1.00	1.00	1.00	1.00	1.00
2	2.02	1.42	1.08	1.04	1.02
12	13.98	4.75	1.46	1.28	1.04
24	28.41	9.92	2.28	1.57	1.14
60	70.06	28.20	4.49	2.92	1.30
120	260.18	85.19	13.25	7.56	2.72
		B: Nonoverlapping versus Implied			
1	1.00	1.00	1.00	1.00	1.00
2	4.19	1.87	1.17	1.07	1.03
12	162.95	38.62	2.96	1.99	1.16
24	718.49	171.81	8.21	3.26	1.60
60	4340.22	1079.07	42.82	14.69	2.49
120	33022.83	7145.40	264.59	84.66	7.55

The table provides a comparison of the relative efficiency of the long-horizon regression coefficient, calculated from a Monte Carlo empirical distribution using 720 observations and 1000 replications. Three methods are used: (i) overlapping observations, (ii) nonoverlapping observations, and (iii) the implied coefficient from short-horizon regressions under AR(1). The estimation involves the regression of a J-period return on a predetermined variable, X_t, which follows a first-order autoregressive process (with corresponding parameter ρ). Panel A provides the ratio between the simulated variance of the nonoverlapping regression coefficient and the simulated variance of overlapping regression coefficients for various values of J and ρ. Panel B provides results similarly for the ratio between the variance of the nonoverlapping regression coefficient and the implied regression coefficient.

small sample efficiency from the use of overlapping observations above and beyond the efficiency implied by the asymptotic theory. Instead, it reflects the even worse small sample properties of using nonoverlapping regressions with only 6 and 12 observations (most probably due to degrees of freedom being used in estimation). To check this, we also performed simulations using a smaller number of observations, e.g., T equal to 240 (not reported in Table 3.2A). For each J, the ratio of variances of the estimators from regressions using nonoverlapping and overlapping data, respectively, were higher with $T = 240$ than with $T = 720$. This suggests that the asymptotic theory breaks down at smaller sample sizes. Of course, most researchers would not perform regressions with six observations. The results in Tables 3.1A and 3.2A imply that using overlapping data (for highly autocorrelated regressors) should not change this practice in any meaningful way.

3.2. Transforming Long-Horizon Regressions—Is It Sensible?

As mentioned in Section 3.1, one criticism of the long-horizon regression approach is that, in calculating asymptotic distributions, the procedure requires estimation of the variance-covariance matrix of the normal equations in the presence of overlapping data. This can lead to poor small-sample properties when the procedure is applied in practice to data. As a result, a number of papers, most notably Hodrick (1992), have posited an alternative method for long-horizon regressions which avoids the overlapping data difficulty (see also

Jegadeesh 1991 and Cochrane 1991a, among others). This alternative method switches the regression from J-period returns on the predetermined variable to a regression of single-period returns on the same predetermined variable aggregated over J-periods. The idea is that this alternative method falls under standard OLS theory, but nevertheless captures the important covariations between returns and the predetermined variable at long horizons.

For purposes of illustration, first consider the case in which the predetermined variable X_t is independent through time; i.e., $\rho = 0$ in (2.2). Hodrick (1992) and others suggest the following alternative regression to (3.1):

$$R_{t+1} = \alpha^h + \beta_J^h \frac{1}{J} \sum_{i=1}^{J} X_{t+i-J} + \epsilon_{t+1}^h(J), \tag{3.4}$$

where R_{t+1} = continuously compounded return from t to $t + 1$,

β_J^h = regression coefficient using the Hodrick methodology,

$\sum_{i=1}^{J} X_{t+i-J}$ = J-period sum of the predetermined variable.

Note that the corresponding OLS coefficients from regression (3.1) and (3.4) can be written as

$$\hat{\beta}_J^{\text{ol}} = \frac{\text{cov}(\Sigma_{i=1}^J R_{t+i}, X_t)}{\text{var}(X_t)}, \tag{3.5}$$

$$\hat{\beta}_J^h = \frac{\text{cov}(R_{t+1}, \Sigma_{i=1}^J X_{t+i-J})}{(1/J)\ \text{var}(\Sigma_{i=1}^J X_{t+i-J})}. \tag{3.6}$$

Both numerators of the estimators, $\hat{\beta}_J^{\text{ol}}$ in (3.5) and $\hat{\beta}_J^h$ in (3.6), converge in probability to the sum of the true autocovariances of R_t and X_{t+i-J}, denoted $C_{R,X}(J)$. Under stationarity, there is no reason to expect the numerator of $\hat{\beta}_J^{\text{ol}}$ to be a "better" estimator than the numerator of $\hat{\beta}_J^h$ (or vice versa) in small samples.

Similarly, if X_t is independent through time, then the denominators of both estimators converge to the same value: the true variance of X_t, denoted σ_x^2. In contrast to the convergence of the numerator, however, we might expect that in small samples the one-period variance estimator behaves "better" than the J-period estimator. This is because for large J we are effectively reducing the sample size and thus the reliability of the variance estimator. Richardson and Stock (1989) provide a formal discussion of this intuition, and show that, under a different asymptotic setting, the J-period variance estimator may not be consistent. Of practical value, their asymptotic setting seems to provide more accurate approximations in small samples for large J.

In terms of the more standard asymptotic theory in this paper, note that the asymptotic variance of the one-period and J-period variance estimators are, respectively, $M_4 - (\sigma_x^2)^2$ and $M_4 + [J(J - 2)(4J - 1)/3J^2]\,(\sigma_x^2)^2$, where M_4 is the fourth central moment of X_t. For example, in the case of normality, the ratio of the variance of the J-period estimator to the one-period estimator is higher, i.e., $\frac{3}{2} + \frac{2}{3}J$. Nevertheless, $\hat{\beta}_J^{\text{ol}}$ and $\hat{\beta}_J^h$ are functions of not only these variance estimators, but also of the estimators for $C_{R,X}(J)$. Hence, although the alternative methodology seemingly uses the less reliable variance estimator, it is important to understand how this estimator interacts with the estimator for $C_{R,X}(J)$. In fact, under the above assumptions, it turns out that the asymptotic variances of the two estimators, $\hat{\beta}_J^{\mathrm{ol}}$ and $\hat{\beta}_J^h$, are the same, namely $J\ \sigma_R^2/\sigma_x^2$. Thus, at least asymptotically, there is no efficiency gain either way.

Unfortunately, as mentioned above, in small samples and for large J, there is reason to believe that this asymptotic justification breaks down (Richardson and Stock 1989). It seems worthwhile exploring this point via a Monte Carlo simulation. Table 3.3 provides a comparison of the distribution of $\hat{\beta}_J^{\mathrm{ol}}$ and $\hat{\beta}_J^h$ for regressions using 720 observations, with values of ρ equal to 0, .95, and .99, and values of J ranging over 12, 24, 60, 120, and 240. Note that the distributions in Table 3.3 have been adjusted by a constant reflecting the true value of the estimator's variance divided by $T - J$.[8] Thus, the empirical distribution of the estimators, $\hat{\beta}_J^{\mathrm{ol}}$ and $\hat{\beta}_J^h$, obtained via the Monte Carlo simulation should reflect a $N(0,1)$ distribution if the asymptotic theory is a good approximation in small samples. As is clear from Table 3.3, this approximation gets progressively worse as we increase the horizon length J.

In terms of the efficiency of the estimators under the null hypothesis, the distribution of the estimator, $\hat{\beta}_J^{\mathrm{ol}}$, is relatively tighter than the corresponding distribution of the alternative estimator, $\hat{\beta}_J^h$, in small samples. This is especially true as J is increased, confirming the intuition above regarding the unreliability of the multiperiod variance estimator for σ_x^2. For example, consider the horizon length $J = 120$. The variance of the estimator from regression (3.4) (i.e., var($\hat{\beta}_J^h$)) is 2.0, 2.3, and 3.2 times the variance of the estimator from regression (3.1) (i.e., var($\hat{\beta}_J^{\mathrm{ol}}$)), for ρ equal to 0, .95, and .99, respectively. For even larger J, the results are even more damaging for the alternative estimator $\hat{\beta}_J^h$. For instance, in the extreme case of $J = 360$, the variance of $\hat{\beta}_J^h$ is 18.1, 24.9, and 63.2 times greater than the variance of $\hat{\beta}_J^{\mathrm{ol}}$, which also suggests that the problem worsens when using highly autocorrelated regressors.

The above intuition and the results in Table 3.3 suggest that researchers should be cautious in adopting regression equation (3.4) in lieu of the regression equation in (3.1). The advantage is that the econometrician can avoid some of the difficulties inherent in calculating test statistics in the presence of serially correlated errors (i.e., induced by overlapping data in (3.1)). However, the disadvantage is that the estimators themselves are less efficient in small samples, and that this efficiency worsens as the horizon length increases. The problems with overlapping data can be rectified if the econometrician is willing to impose structure on the errors, as in Richardson and Smith (1991) or in Hodrick's (1992) extension to multivariate data (though this may depend on the econometrician's interest in alternative

[8] For any ρ, the transformed regression is actually

$$R_{t+1} = \alpha^h + \beta_J^h \frac{1}{f(\rho, J)} \sum_{i=1}^{J} X_{t+i-J} + \epsilon_{t+1}^h (J),$$

where

$$f(\rho, J) = \left[J + \frac{2\rho}{1-\rho} \left(J - 1 - \rho \frac{1-\rho^{J-1}}{1-\rho} \right) \right].$$

The regression is transformed by a constant in this way so as to make the regression coefficients from the two methods comparable.

TABLE 3.3
Comparison of Long-Horizon Regression Specifications

Regression	J	ρ	Percentiles .05	.10	.50	.9	.95	Var. of Reg. Coeff. Estimated	Asymptotic	Ratio of Var.
J on 1	12	0.000	−1.644	−1.262	0.033	1.292	1.669	0.017	0.017	
1 on J	12	0.000	−1.655	−1.273	0.032	1.280	1.607	0.017	0.017	1.006
J on 1	12	0.950	−1.509	−1.180	0.080	1.369	1.777	0.177	0.168	
1 on J	12	0.950	−1.646	−1.233	0.061	1.417	1.862	0.190	0.168	1.074
J on 1	12	0.990	−1.974	−1.579	0.015	1.503	2.004	0.303	0.196	
1 on J	12	0.990	−2.225	−1.668	0.067	1.531	2.003	0.334	0.196	1.103
J on 1	24	0.000	−1.640	−1.220	0.039	1.262	1.622	0.034	0.034	
1 on J	24	0.000	−1.766	−1.417	0.003	1.312	1.774	0.040	0.034	1.167
J on 1	24	0.950	−1.748	−1.351	−0.015	1.352	1.660	0.597	0.572	
1 on J	24	0.950	−1.706	−1.322	−0.010	1.276	1.606	0.611	0.572	1.023
J on 1	24	0.990	−1.819	−1.468	0.048	1.489	2.093	1.087	0.765	
1 on J	24	0.990	−2.155	−1.557	0.096	1.803	2.328	1.404	0.765	1.292
J on 1	60	0.000	−1.554	−1.242	0.027	1.304	1.641	0.089	0.091	
1 on J	60	0.000	−1.899	−1.448	0.005	1.417	1.769	0.111	0.091	1.247
J on 1	60	0.950	−1.488	−1.131	−0.008	1.214	1.595	2.140	2.447	
1 on J	60	0.950	−1.831	−1.329	0.109	1.497	1.987	3.323	2.447	1.553
J on 1	60	0.990	−1.945	−1.460	−0.025	1.426	1.921	6.115	4.506	
1 on J	60	0.990	−2.232	−1.695	−0.031	1.699	2.379	9.586	4.506	1.568
J on 1	120	0.000	−1.414	−1.047	0.013	1.099	1.503	0.154	0.200	
1 on J	120	0.000	−2.077	−1.528	−0.054	1.398	1.808	0.302	0.200	1.967
J on 1	120	0.950	−1.473	−1.149	−0.012	1.126	1.462	5.318	6.536	
1 on J	120	0.950	−2.218	−1.620	0.068	1.607	2.172	12.285	6.536	2.310
J on 1	120	0.990	−1.736	−1.351	−0.012	1.262	1.652	17.805	16.680	
1 on J	120	0.990	−2.841	−1.959	−0.009	2.138	2.912	56.715	16.680	3.185
J on 1	240	0.000	−1.250	−0.953	−0.004	0.867	1.123	0.288	0.500	
1 on J	240	0.000	−2.856	−2.039	−0.062	1.933	2.624	1.438	0.500	4.991
J on 1	240	0.950	−1.181	−0.895	−0.013	0.890	1.275	9.891	17.917	
1 on J	240	0.950	−3.177	−2.164	0.107	2.324	3.215	71.595	17.917	7.239
J on 1	240	0.990	−1.346	−1.013	−0.016	1.015	1.320	40.642	61.947	
1 on J	240	0.990	−4.420	−2.646	0.031	2.764	3.841	416.605	61.947	10.251
J on 1	360	0.000	−0.887	−0.603	−0.000	0.697	0.917	0.311	1.000	
1 on J	360	0.000	−3.711	−2.681	0.091	2.634	3.859	5.632	1.000	18.107
J on 1	360	0.950	−0.863	−0.642	0.030	0.662	0.916	10.900	36.889	
1 on J	360	0.950	−4.403	−3.110	−0.003	3.009	4.388	271.420	36.889	24.901
J on 1	360	0.990	−0.901	−0.654	−0.031	0.640	0.835	40.057	145.476	
1 on J	360	0.990	−6.199	−3.890	0.025	4.082	6.236	2530.852	145.476	63.181

The table provides a comparison of the relative efficiency of long-horizon regression coefficients estimated two different ways: (i) a regression of J-period returns on a single-period X_t (denote J on 1), and (ii) a regression of single-period returns on a J-period sum of X_t (denote 1 on J). The results are simulated using 1000 replications of 720 observations on an i.i.d. return process and an AR(1) process on X_t (with corresponding AR(1) parameter ρ). The table provides both the empirical distribution of the regression coefficients and also the ratio of the simulated variances of the coefficients. Note that the estimators are equivalent in large samples in every respect, including their asymptotic distribution. The table has adjusted their small-sample distributions by their true asymptotic variance to approximate the implied standard normal distribution of the estimators.

theories to the null hypothesis; e.g., see Campbell 1993b). Unfortunately, there is no cure for the poor efficiency of the alternative estimator, $\hat{\beta}_J^h$. Its problem derives from having to estimate the multiperiod variance of X_t, which is the estimator's central feature.

3.3 A Comparison of Long-Horizon versus Short-Horizon Regressions

An alternative approach to estimating coefficients from long-horizon regressions is to derive implied long-horizon coefficients from short-horizon regressions. This approach has been adopted by Campbell and Shiller (1987), Kandel and Stambaugh (1988), Campbell (1993a), Hodrick (1992), and Bekaert and Hodrick (1992), among others. The intuition underlying their estimation approach is straightforward. It is based on the fact that estimating short-horizon systems is generally more efficient than estimating long-horizon systems. Thus, using these more efficient short-horizon estimators to estimate long-horizon coefficients should also be more efficient. Below, we quantify this efficiency gain in the context of the statistical setting used in this paper.

For purposes of illustration, consider the regression system

$$R_{t+1} = \alpha_1 + \beta_1 X_t + \epsilon_{t+1}^h (1), \tag{3.7}$$

$$X_{t+1} = \mu + \rho X_t + \eta_{t+1}. \tag{3.8}$$

Under these assumptions, it is possible to derive an implied counterpart to the regression coefficient β_J estimated in regressions (2.1), (3.1), and (3.4). In particular, denote the OLS regression coefficients from (3.7) and (3.8) as $\hat{\beta}_1$ and $\hat{\rho}$. The implied consistent estimator of β_J, as a function of the two one-period estimators ($\hat{\beta}_1$, and $\hat{\rho}$), denoted β_J^{imp}, can be written as

$$\hat{\beta}_J^{\text{imp}} = \hat{\beta}_1 \frac{1 - \hat{\rho}^J}{1 - \hat{\rho}}. \tag{3.9}$$

The GMM procedure described in Section 3.1 can be used to estimate the variance of this estimator. Specifically, under the null hypothesis $\beta_J = 0$, it is possible to show that

$$\text{var}(\hat{\beta}_J^{\text{imp}}) = \frac{\sigma_R^2}{\sigma_x^2} \left(\frac{1 - \rho^J}{1 - \rho} \right)^2. \tag{3.10}$$

Similar to the comparison given in Section 3.1, we can look at the relative asymptotic efficiency of the estimators, $\hat{\beta}_J^{\text{nol}}$, $\hat{\beta}_J^{\text{ol}}$, and $\hat{\beta}_J^{\text{imp}}$. Consistent with intuition, the short-horizon coefficient is always more efficient. However, this efficiency gets drastically reduced for highly autocorrelated X_t.

Table 3.1B provides calculations of the ratio of asymptotic variances given in (3.3) and (3.10) for a variety of J and ρ. In particular, we vary ρ over the range (0, .5, .9, .95, .99) and J over the range (2, 12, 24, 60, 120). As an example, consider the estimator's relative efficiency for $\rho = 0$ versus $\rho = .99$. Recall that for $\rho = 0$, the ratios of the variances of $\hat{\beta}_J^{\text{nol}}$ and $\hat{\beta}_J^{\text{ol}}$ are 2, 12, 24, 60, and 120 when J equals 2, 12, 24, 60, and 120 respectively. When comparing $\hat{\beta}_J^{\text{nol}}$ to $\hat{\beta}_J^{\text{imp}}$, the ratio increases to 4, 144, 576, 3600, and 14400 respectively. That is, consider using 720 monthly observations to forecast five-year returns. For this case in which the forecasting variable, X_t, is uncorrelated through time, then the implied long-horizon approach would be as efficient as a nonoverlapping regression which uses 43,200 five-year independent observations, or roughly 216,000 years of data!

In contrast, for $\rho = .99$, recall that the ratios of the variances of $\hat{\beta}_J^{\text{nol}}$ and $\hat{\beta}_J^{\text{ol}}$ are 1.02, 1.04, 1.14, 1.30, and 2.72 when J equals 2, 12, 24, 60, and 120 respectively. When comparing $\hat{\beta}_J^{\text{nol}}$ to $\hat{\beta}_J^{\text{imp}}$, the ratio increases to only 1.03, 1.11, 1.32, 1.89, and 5.55 respectively. That is, in the above example, the implied long-horizon approach would be as efficient as a nonoverlapping regression which uses 23 five-year independent observations, or a little less than twice the initial data. At least in terms of this example, the implied long-horizon approach seems preferable to the more standard long-horizon regression. Nevertheless, the overall conclusion from Table 3.1B is that, for highly autocorrelated regressors, the gains above and beyond the nonoverlapping method are relatively small.

These analytical results given in Table 3.1B are strictly valid only asymptotically. It seems worthwhile exploring whether the intuition carries through to small samples. Using the Monte Carlo simulation described in Section 2, we compare the small sample variances and Monte Carlo empirical distribution of $\hat{\beta}_J^{\text{nol}}$, $\hat{\beta}_J^{\text{ol}}$, and $\hat{\beta}_J^{\text{imp}}$. In particular, to coincide with existing studies, we choose the number of observations T equal to 720; the level of the autocorrelation parameter ρ to vary over the range (0, .5, .9, .95, .99); and the horizon length J to vary over the range (2, 12, 24, 60, 120). The results for these simulations are provided in Table 3.2B.

Several features of the simulation evidence in Table 3.2B are especially interesting. First, the small sample simulation results are similar in spirit to the analytical results in Table 3.1B. For example, the relative efficiency of the short-horizon method increases with the horizon length J, but declines rapidly for large ρ. The magnitudes are also consistent with those in Table 3.1B. Second, the short-horizon approach is for the most part more efficient than the long-horizon regression methods. For example, in every simulation, the ratios of var($\hat{\beta}_J^{\text{ol}}$) and var($\hat{\beta}_J^{\text{nol}}$) to var($\hat{\beta}_J^{\text{imp}}$) are greater than one. Third, this efficiency, however, is greatly reduced for highly autocorrelated regressors. For instance, consider setting the horizon length J equal to 60. In this case, the ratios of var($\hat{\beta}_J^{\text{ol}}$) and var($\hat{\beta}_J^{\text{nol}}$) to the implied long-horizon estimator's variance are, respectively, 61.9, 38.3, 9.5, 5.0, and 1.9 and 4340.2, 1079.1, 42.8, 14.7, and 2.5 for ρ equal to 0, .5, .9, .95, and .99.

Since we know that the regressors tend to be highly autocorrelated, and that a regression with large J provides very few independent observations (e.g., see Section 3.1), this suggests that the short-horizon approach, even with its enhanced efficiency, is not a solution to the long-horizon regression problem.

This can be further illustrated by recognizing that there is a potential cost to using the short-horizon approach for inferring long-horizon coefficients. Suppose the true process on X_t is not an AR(1), but instead an AR(2); i.e.,

$$X_{t+1} = \mu + \delta_1 X_t + \delta_2 X_{t-1} + \eta_t.$$

Under this alternative model for X_t, it is possible to show that the implied long-horizon coefficient estimator converges to

$$\text{plim}\ (\hat{\beta}_J^{\text{imp}}) = \beta_1 \frac{(1 - \delta_2)^J - \delta_1^J}{(1 - \delta_2 - \delta_1)(1 - \delta_2)^{J-1}}. \tag{3.11}$$

Clearly, if $\beta_1 \neq 0$ (i.e., the alternative is true), then $\hat{\beta}_J^{\text{imp}}$ will not be a consistent estimator for the true long-horizon coefficient β_J^{ol}. To see this, note that in general

$$(3.12) \qquad \beta_J^{\text{ol}} = \beta_1[1 + \sum_{i=1}^{J-1} \text{corr}(X_t, X_{t-i})],$$

where corr(·) denotes the correlation between two variables. However, in the AR(2) example, the jth-order autocorrelations of X_t are given by $\text{corr}(X_t, X_{t-1}) = \delta_1/(1 - \delta_2)$, $\text{corr}(X_t, X_{t-2}) = \delta_2 + \delta_1^2/(1 - \delta_2)$, and so on. If $\delta_2 \neq 0$, then there is no way for (3.11) and (3.12) to coincide; that is, $\text{plim}(\hat{\beta}_J^{\text{imp}} \neq \text{plim}(\beta_J^{\text{ol}})$.

In contrast, the long-horizon regression coefficient $\hat{\beta}_J^{\text{ol}}$ is always a consistent estimator for β_J under both the null $\beta_1 = 0$ and alternative hypotheses $\beta_1 \neq 0$. This condition is especially important because financial economists would like to interpret the predictability if it shows up in the data. Thus, if we are to interpret the point estimates of the coefficient β_J in any meaningful way, consistency seems like a minimum criterion. If we use the short-horizon approach, however, consistency is lost if the autoregression is underspecified.

A possible solution to this consistency problem is to overspecify the autoregressive structure of the forecasting variable, X_t. While this does not pose a problem asymptotically under the null $\beta_J = 0$, it is conceivable that overfitting could lead to difficulties in small samples. As a check, we performed some Monte Carlo tests in a simple overfitting example. Specifically, we treated the X_t as an AR(2), and calculated the implied $\hat{\beta}_J^{\text{imp}}$ using simulated data in which X_t follows an AR(1). For this particular example, overfitting had no effect on the results.[9]

4. RELATED WORK AND FUTURE RESEARCH

This paper takes as given the long-horizon regression problem. It is worthwhile commenting on the motivation for using long-horizon regressions in studying the predictability of stock returns. From an economic point of view, there is general interest in the magnitude of long-horizon coefficients, and long-horizon regressions can be justified from that perspective. From a purely statistical point of view, Campbell (1993b) argues that long-horizon regressions may be more powerful than a single-horizon regression even if the system is completely specified as in Section 2 (see also Cochrane 1991a). This claim seems to be at odds with standard econometric theory, which would imply one can do no better than the system of equations described in (3.7) and (3.8). Stock (1992) provides a response to an earlier version of Campbell's paper, and argues that this increased power is just an anomaly of highly autoregressive predetermined variables. Furthermore, in terms of practical use, he argues that the size of long-horizon tests are so poor that these anomalies cannot be taken advantage of in actual applications of the data. Campbell (1993b) performs some Monte Carlo simulations, which leaves the debate open.

While this literature is still developing, an intriguing argument for long-horizon regressions has been given by Stambaugh (1993). Stambaugh suggests that violations of OLS, such as heteroskedasticity, can make long-horizon regressions more effective. Specifically, in a simple GARCH framework, Stambaugh reports efficiency gains by going to a long-

[9]It should be pointed out that, in practice, substantial overspecification of the AR process could have an effect in small samples (see Box and Jenkins 1976). Thus, the econometrician needs to treat each estimation problem on a case-by-case basis.

horizon setting. Stambaugh's work suggests that long-horizon regressions will remain important in financial applications.

With respect to the impact of heteroskedasticity on the efficiency of the various procedures for estimating β_J, Stambaugh (1993) finds that the presence of heteroskedasticity leads to even greater relative efficiency of the overlapping to the nonoverlapping case. Although the increase tends to be small, it is magnified for highly autocorrelated predetermined variables, as well as persistent variances of returns (both of which are applicable to financial data). Of more interest, in his heteroskedasticity-consistent framework, Stambaugh shows that the implied long-horizon approach may actually be less efficient than the long-horizon regression method, a reversal of the results reported in this paper. This suggests that researchers should be especially aware of the distributional structure of the asset returns they are forecasting.

Note that the results in this paper have also been derived under the null hypothesis of no predictability. Stambaugh (1993) finds that the gains in efficiency over the nonoverlapping approach are slightly magnified for nonzero β_J alternatives. In contrast, Campbell (1993b) reports much larger gains than under the null $\beta_J = 0$. This is a result of his specification of how expected returns evolve through time. In general though, the results in these papers coincide with the basic finding here that the persistence of X_t is the most important factor determining the efficiency of the various long-horizon regression methods.

On a more negative note, for large J, it makes little sense to use long-horizon regressions with nonoverlapping observations. For example, with 60 years of data, five-year return regressions leave the econometrician with only 12 observations. Unfortunately, for the case of highly autocorrelated predetermined variables, the results in this paper imply that the other estimation approaches are not reasonable alternative procedures. This result regarding the efficiency loss due to correlation may be endemic of many other applications in finance. A recent paper by Graham (1993) quantifies this loss and justifies the importance of correlation as a tool in understanding the relation between information content and the number of data observations.

REFERENCES

BAHADUR, R. R. (1960): "Stochastic Comparison of Tests," *Ann. Math. Statist.* 31, 276–295.

BEKAERT, G., and R. J. HODRICK (1992): "Characterizing Predictable Components in Equity and Foreign Exchange Rates of Return," *J. Finance,* 47, 467–509.

BOUDOUKH, J., and M. P. RICHARDSON (1993): "Stock Returns and Inflation: A Long-Run Perspective," *Am. Econ. Rev.,* forthcoming.

BOX, G. E., and G. M. JENKINS (1976): *Time Series Analysis: Forecasting and Control.* Oakland, CA: Holden-Day.

CAMPBELL, J. (1993a): "A Variance Decomposition of Stock Returns," *Econ. J.,* 101, 152–179.

CAMPBELL, J. (1993b): "Why Long Horizons? A Study of Power against Persistent Alternatives," working paper, Princeton University.

CAMPBELL, J., and R. J. SHILLER (1987): "Cointegration and Tests of Present Value Models," *J. Political Econ.,* 95, 1062–1088.

COCHRANE, J. H. (1991a): "Volatility Tests and Efficient Markets: A Review Essay," *J. Monetary Econ.,* 27, 463–485.

COCHRANE, J. H. (1991b): "Production-Based Asset Pricing and the Link between Stock Returns and Economic Fluctuations," *J. Finance,* 46, 207–234.

FAMA, E. F., and R. R. BLISS (1987): "The Information in Long-Maturity Forward Rates," *Am. Econ. Rev.,* 77, 680–692.

FAMA, E. F., and K. R. FRENCH (1988): "Dividend Yields and Expected Stock Returns," *J. Financial Econ.*, 22, 3–25.

FAMA, E. F., and K. R. FRENCH (1989): "Business Conditions and Expected Returns on Stocks and Bonds," *J. Financial Econ.*, 25, 3–26.

FOSTER, F. D., and T. SMITH (1992): "Assessing Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," working paper, Fuqua School of Business, Duke University.

GEWEKE, J. (1981): "The Approximate Slope of Econometric Tests," *Econometrica*, 49, 1427–1442.

GOETZMANN, W. N., and P. JORION (1993): "Testing the Predictive Power of Dividend Yields," *J. Finance*, 48, 663–679.

GRAHAM, JOHN R. (1993): "Measuring the Information Loss Attributable to Correlation," mimeo, Fuqua School of Business, Duke University.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L. P., and R. J. HODRICK (1980): "Forward Rates as Predictors of Future Spot Rates: An Econometric Analysis," *J. Political Econ.*, 88, 829–853.

HODRICK, R. J. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Rev. Financial Stud.*, 5, 357–386.

JEGADEESH, N. (1991): "Seasonality in Stock Price Mean Reversion: Evidence from the U.S. and U.K.," *J. Finance*, 46, 1427–1444.

KANDEL, S., and STAMBAUGH, R. F. (1988): "Modeling Expected Stock Returns for Long and Short Horizons," working paper, Wharton School, University of Pennsylvania.

NELSON, C., and M. J. KIM (1993): "Predictable Stock Returns: The Role of Small Sample Bias," *J. Finance*, 48, 641–661.

RICHARDSON, M., and T. SMITH (1991): "Tests of Financial Models in the Presence of Overlapping Observations," *Rev. Financial Stud.*, 4, 227–254.

RICHARDSON, M., and T. SMITH (1992): "A Unified Approach to Testing for Serial Correlation in Stock Returns," *J. Business*, forthcoming.

RICHARDSON, M., and J. H. STOCK (1989): "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *J. Financial Econ.*, 25, 323–348.

STAMBAUGH, R. F. (1993): "Estimating Conditional Expectations when Volatility Fluctuates," working paper, Wharton School, University of Pennsylvania.

STOCK, J. H. (1992): Discussion, National Bureau of Economic Research Asset Pricing Conference, University of Pennsylvania.

WHITELAW, R. F. (1993): "Time Variations and Covariations in the Expectation and Volatility of Stock Market Returns," *J. Finance*, forthcoming.