Measuring the Predictable Variation in Stock and Bond Returns

Chris Kirby
Rice University
University of Texas at Dallas

Recent studies show that when a regression model is used to forecast stock and bond returns, the sample R^2 increases dramatically with the length of the return horizon. These studies argue, therefore, that long-horizon returns are highly predictable. This article presents evidence that suggests otherwise. Long-horizon regressions can easily yield large values of the sample R^2, even if the population R^2 is small or zero. Moreover, long-horizon regressions with a small or zero population R^2 can produce t-ratios that might be interpreted as evidence of strong predictability. In general, the analysis provides little support for the view that long-horizon returns are highly predictable.

The theory that stock and bond markets process information efficiently is a hallmark of modern finance. Although the traditional formulation of the efficient markets hypothesis rules out the ability to predict returns, most financial economists would probably agree that this view of efficiency is outdated. Recent

---

A previous version of this article was circulated under the title "A Multivariate Analysis of the Predictable Variation in Stock and Bond Returns." The current version is based on ideas developed in my dissertation at Duke University. I thank the members of my committee—Cam Harvey, David Hsieh, Tom Smith, George Tauchen, and Bob Whaley—for providing valuable support and suggestions. I also appreciate the comments and suggestions of Phillip Braun, Doug Foster, Andy Lo (the editor), Barbara Ostdiek, an anonymous referee, seminar participants at Duke University, and session participants at the 1993 meetings of the Western Finance Association. Part of this research was completed while the author was visiting the University of Michigan and the University of Maryland. Address correspondence to Chris Kirby, School of Management, University of Texas at Dallas, Richardson, TX 75083-0688.

The Review of Financial Studies Fall 1997 Vol. 10, No. 3, pp. 579–630

advances in asset pricing theory, along with mounting empirical evidence of predictability, have persuaded the majority of researchers to abandon the constant expected returns paradigm. Nevertheless, certain aspects of the empirical research on predicting returns remain controversial. A number of studies, for example, report that stock and bond returns appear to exhibit a striking degree of predictability over long horizons. This evidence is not necessarily inconsistent with the view that markets are efficient, but it does seem to contradict much of the conventional wisdom in this regard.

Almost all of the research on predicting long-horizon returns falls under the general heading of regression analysis. The basic strategy adopted in most studies is to regress overlapping returns for various holding periods on a set of predetermined instrumental variables. In the majority of cases, the authors of such studies treat the sample R^2 from the regression specification as a measure of the economic significance of the predictable component of returns. Fama and French (1988a), for example, argue that dividend yields explain a large fraction of the total variation in long-horizon stock returns. To support this claim they show that the sample R^2 increases from around 3% for monthly returns to well over 25% for four-year returns. Campbell and Shiller (1988) cite similar evidence in their study of the link between dividend yields, earnings:price ratios, and stock returns. They report large values of the sample R^2 for both 3-year and 10-year returns and conclude, like Fama and French (1988a), that long-horizon stock returns are indeed highly predictable.

The apparent pattern of strong predictability at long horizons extends to other classes of assets as well. In a follow-up to their initial study, Fama and French (1989) demonstrate that two interest rate variables—a term spread and a default-risk spread—seem to explain a substantial fraction of the long-term variation in bond returns. Again they observe a dramatic increase in the sample R^2 as the return horizon grows from one month to four years. The sample R^2 is usually less than 10% for monthly and quarterly bond returns, but often exceeds 30% for returns measured over longer horizons. Fama and French attribute this increase in explanatory power at long horizons to low-frequency oscillations in expected returns. They further contend that these low-frequency oscillations reflect the rational response of investors to slowly changing business conditions.

This tendency to treat the sample R^2 as a measure of the economic significance of predictability is not surprising. After all, the approach does have a substantial degree of intuitive appeal. The important point to remember, however, is that the least-squares theory used to justify this practice rests largely on the assumption that the error term for the regression model can be treated as an independently and identically distributed normal random variable. Outside the context of the classical least-squares setting, the distributional properties of the sample R^2 are largely unknown and its relation to formal tests for predictability is far from clear. It seems, therefore, that the authors of some of the most heavily cited studies of long-horizon predictability overlook an important consideration. Large values of the sample R^2 might actually be consistent with the hypothesis that long-horizon stock and bond returns are either unpredictable or only slightly predictable.

The idea that large values of the sample R^2 can be statistically insignificant is not new. Granger and Newbold (1974) perform simulations where both the dependent and independent variables in a regression model follow a random walk. Their simulations often yield large values of the sample R^2 even though the population value is zero by construction. In a more recent study, Goetzmann and Jorion (1993) question the evidence that dividend yields have the power to forecast stock returns. They use a Monte Carlo approach to examine the empirical distribution of various regression criteria under the null hypothesis that asset returns are unpredictable. In simulations with data similar to that of Fama and French (1988a), the mean value of the sample R^2 increases from less than 1% for monthly returns to around 12% for four-year returns. As a result, Goetzmann and Jorion argue that there is little evidence to indicate that dividend yields can be used to forecast stock returns.

This article offers a new perspective on the evidence that stock and bond returns are highly predictable over long horizons. The analysis draws on standard asymptotic arguments to derive the limiting distribution of the sample R^2 under both the null hypothesis that returns are unpredictable and the alternative hypothesis that returns contain a predictable component. Once the features of this distribution are established, it becomes clear that serial correlation and conditional heteroscedasticity can have a pronounced effect on the sampling properties of the R^2 statistic. The key feature of the distribution can be summarized as follows: serial correlation, including the type induced by the use of overlapping returns, generally leads to a substantial increase in the standard error of the sample R^2. Thus, large realizations of the sample R^2 may not, in and of themselves, indicate that long-horizon stock and bond returns are highly predictable.

The empirical investigation confirms that this is the case. Although the sample R^2 does increase with the length of the return horizon, the associated standard errors reveal that the long-horizon results are quite imprecise. In general, the inference that long-horizon returns are highly predictable does not appear to be justified. Evidence from a Monte Carlo experiment lends additional credence to the empirical findings. The t-ratios for the slope coefficients are found to reject the null hypothesis far too often in samples of the size typically used in long-horizon studies. Moreover, regression models with a small population R^2 can produce t-ratios that might be interpreted as evidence of strong predictability. All of the results point to a simple conclusion: it is vital to consider the distributional properties of the t-ratios and sample R^2 when drawing inferences about the ability to predict long-horizon returns.

1. Asymptotic Theory for Long-Horizon Regressions

Over the years, academics and practitioners alike have devoted a great deal of effort to the search for predictable variation in stock and bond returns [see, for example, Fama and Schwert (1977), Rozeff (1984), Keim and Stambaugh (1986), Campbell (1987), Fama and Bliss (1987), Campbell and Shiller (1988), Fama and French (1988a, b, 1989), Jegadeesh (1991), Ferson and Harvey (1991), and Bekaert and Hodrick (1992)]. Although the bulk of the empirical research on forecasting returns uses monthly data, a review of the recent literature reveals an increasing number of studies that focus on predictability over longer horizons. These long-horizon studies typically have a number of characteristics in common: (i) they use sample sizes that are relatively small; (ii) they assess the ability to predict returns based on a regression specification that, more often than not, uses overlapping returns; and (iii) they treat the sample R^2 from the regression model as a measure of the economic significance of the predictable component of returns. It is important, therefore, to understand how the use of small sample sizes and overlapping returns affects the distributional properties of the sample R^2. The analysis begins with an overview of the asymptotic theory for dynamic linear models. All proofs appear in the Appendix.

1.1 Asymptotic normality of the slope coefficients

There are a number of different ways to assess whether asset returns contain a predictable component. As a general rule, however, researchers tend to use some type of regression specification. The most common approach is to estimate a multiple regression model of the form

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \tilde{z}_t' \beta_k + \tilde{\varepsilon}_{t+k}, \tag{1}$$

where $\sum_{i=1}^{k} \tilde{r}_{t+i}$ is the k-period excess return on a portfolio of assets, $\tilde{z}_t$ is the date t realization of a $m \times 1$ vector of instrumental variables, and $\tilde{\varepsilon}_{t+k}$ is a mean-zero error term. Because the regressors for the model shown in Equation (1) are stochastic, the least-squares estimator of the vector of slope coefficients is given by

$$\hat{\boldsymbol{\beta}}_k \equiv \hat{\boldsymbol{\Sigma}}_{zz}^{-1} \hat{\boldsymbol{\sigma}}_{zk}, \tag{2}$$

where $\hat{\boldsymbol{\Sigma}}_{zz}$ denotes the sample analog of the variance-covariance matrix of the instruments, and $\hat{\boldsymbol{\sigma}}_{zk}$ denotes the $m \times 1$ vector of sample covariances. The limiting distribution of the least-squares estimator is provided by the following theorem.

Theorem 1.1. *Let $\{\tilde{r}_t, \tilde{\mathbf{z}}_t\}$ be a stationary and ergodic process. Further assume that the regularity conditions given by Hansen (1982) are satisfied. Then the limiting distribution of the least-squares estimator of $\boldsymbol{\beta}_k$ is*

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}_k) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}), \tag{3}$$

where T denotes the number of observations in the data set. The variance-covariance matrix $\boldsymbol{V}$ is given by

$$\boldsymbol{V} \equiv \boldsymbol{\Sigma}_{zz}^{-1} \left[\sum_{j=-\infty}^{\infty} E\left[\tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j}(\tilde{\mathbf{z}}_t - \boldsymbol{\mu}_z)(\tilde{\mathbf{z}}_{t-j} - \boldsymbol{\mu}_z)'\right] \right] \boldsymbol{\Sigma}_{zz}^{-1}, \tag{4}$$

where $\tilde{\varepsilon}_{t+k-j} \equiv (\sum_{i=1}^{k} \tilde{r}_{t+i-j} - \alpha_k - \tilde{\mathbf{z}}'_{t-j}\boldsymbol{\beta}_k)$, and $\boldsymbol{\mu}_z$ denotes the expected value of the vector $\tilde{\mathbf{z}}_t$.

It certainly comes as no surprise that the limiting distribution of the least-squares estimator is multivariate normal. But one feature of the distribution in Equation (3) deserves close attention. Note how the autocovariance structure of the disturbance vector $\tilde{\boldsymbol{\eta}}_{t+k} \equiv \tilde{\varepsilon}_{t+k}(\tilde{\mathbf{z}}_t - \boldsymbol{\mu}_z)$ affects the variance-covariance matrix of the estimator. This relation is important because the matrix $\boldsymbol{V}$ plays a large role in determining the limiting distribution of the sample R^2. To see why, recall that the sample R^2 can be written as a quadratic form in the least-squares estimator of the vector of slope coefficients.[1] If serial correlation or heteroscedasticity make it impossible to obtain a precise estimate of $\boldsymbol{\beta}_k$, then the sample R^2 will reflect a similar degree of imprecision. Thus, the statistical properties of the disturbance vector $\tilde{\boldsymbol{\eta}}_{t+k}$ have a

[1] The sample R^2 is equal to the sample variance of the fitted values divided by the sample variance of the dependent variable. Applying this definition to the model shown in Equation (1) yields:

$$R_k^2 = \left(\frac{\hat{\boldsymbol{\beta}}_k' \hat{\boldsymbol{\Sigma}}_{zz} \hat{\boldsymbol{\beta}}_k}{\hat{\sigma}_k^2} \right),$$

where $\hat{\sigma}_k^2$ denotes the sample variance of the k-period excess return on the portfolio.

great deal to do with the distributional properties of the sample R^2. One of the goals of this article is to develop formal theoretical results that clearly illustrate this relation. A natural first step in this process is to consider the case where the assumptions of classical regression analysis are satisfied.

1.2 Serially uncorrelated homoscedastic disturbances

Let the null hypothesis H_0 be that asset returns are unpredictable. In addition, assume that the disturbance vector $\tilde{\boldsymbol{\eta}}_{t+k}$ is serially uncorrelated and that $\tilde{\varepsilon}_{t+k}$ is conditionally homoscedastic.[2] Within the context of the regression model shown in Equation (1), the testable implication of the null is that $\boldsymbol{\beta}_k = \mathbf{0}$. When we impose the null hypothesis and incorporate the stated assumptions, the limiting distribution of the sample R^2 takes the form given in the following theorem.

Theorem 1.2. *Let $\{\tilde{r}_t, \tilde{z}_t\}$ be a stationary and ergodic process. Further assume that the regularity conditions of Hansen (1982) are satisfied, that $cov(\tilde{\boldsymbol{\eta}}_{t+k}, \tilde{\boldsymbol{\eta}}_{t+k-\tau}) = \mathbf{0}$ for all $\tau \neq 0$, and that $var(\tilde{\boldsymbol{\eta}}_{t+k}) = \sigma_\varepsilon^2 \boldsymbol{\Sigma}_{zz}$, where σ_ε^2 denotes the variance of $\tilde{\varepsilon}_{t+k}$. Then, under the null hypothesis H_0, the limiting distribution of the sample R^2 for the model shown in Equation (1) is*

$$TR_k^2 \xrightarrow{d} \chi_m^2, \tag{5}$$

where χ_m^2 denotes a central chi-square distribution with m degrees of freedom.

It is not difficult to see why the quantity TR_k^2 is asymptotically distributed as a chi-square random variable. Under the conditions stated in the theorem, the matrix $\boldsymbol{V}$ in Equation (4) reduces to $\sigma_k^2 \boldsymbol{\Sigma}_{zz}^{-1}$, where σ_k^2 denotes the variance of the k-period excess return. As a result, the Wald statistic for testing the null hypothesis is

$$\hat{W}_D \equiv T\left(\frac{\hat{\boldsymbol{\beta}}_k' \hat{\boldsymbol{\Sigma}}_{zz} \hat{\boldsymbol{\beta}}_k}{\hat{\sigma}_k^2}\right), \tag{6}$$

where $\hat{\sigma}_k^2$ denotes the sample analog of σ_k^2. Once the ratio inside the parentheses is recognized as the sample R^2 for the regression model of Equation (1), the distributional results of Theorem 1.2 follow immediately.[3]

---

[2] One scenario consistent with this assumption is that $\tilde{z}_t$ is serially uncorrelated and distributed independently of $\tilde{r}_{t+i}$ for all i.

[3] It is well known that, under the null hypothesis, the Wald statistic in Equation (6) converges in distribution to a chi-square random variable with m degrees of freedom.

The asymptotic results shown in Equations (5) and (6) bear a distinct resemblance to the small sample results for the multivariate normal regression model. This similarity stems in large part from the assumption that (i) $\tilde{\eta}_{t+k}$ is serially uncorrelated; and (ii) the error term for the regression model is conditionally homoscedastic. In situations where these assumptions are plausible, the sample R^2 represents an appealing criterion for evaluating whether the null hypothesis of no predictability is credible. To obtain significance points for the sample R^2, we simply divide the values for the appropriate chi-square distribution by the number of observations used in the analysis. If the observed value of the sample R^2 exceeds the cutoff, then the null hypothesis that asset returns are unpredictable is rejected. It is important to keep in mind, of course, that tests performed in this manner are valid only if the aforementioned restrictions on $\tilde{\eta}_{t+k}$ and $\tilde{\varepsilon}_{t+k}$ are satisfied. The effect of relaxing these restrictions is considered in Section 1.3.

1.3 Autocorrelated heteroscedastic disturbances

Stock and bond returns are known to exhibit a marked degree of conditional heteroscedasticity, and overlapping returns are autocorrelated by construction. As a consequence, the sampling theory for long-horizon models must be able to accommodate both serial correlation and conditional heteroscedasticity of unknown form. If the previous analysis is modified to permit the sorts of intertemporal dependence and conditional heterogeneity that may exist in data generated by a stationary and ergodic process, then the limiting distribution of the sample R^2 under null hypothesis H_0 takes the form given in Theorem 1.3.

Theorem 1.3. *Let $\{\tilde{r}_t, \tilde{z}_t\}$ be a stationary and ergodic process. Further assume that the regularity conditions given by Hansen (1982) are satisfied. In addition, allow $\tilde{\eta}_{t+k}$ and $\tilde{\varepsilon}_{t+k}$ to exhibit the type of serial correlation and conditional heteroscedasticity that is consistent with data generated by a stationary and ergodic process. Then, under the null hypothesis H_0, the limiting distribution of the sample R^2 for the model shown in Equation (1) is*

$$TR_k^2 \xrightarrow{d} Q, \tag{7}$$

where Q denotes the general distribution of a quadratic form in a multivariate normal random vector. The mean and variance of Q are

$$\mu_q \equiv tr\,(\sigma_k^{-2}\mathbf{V}\boldsymbol{\Sigma}_{zz}) \qquad and \qquad \sigma_q^2 \equiv 2\,tr\,(\sigma_k^{-2}\mathbf{V}\boldsymbol{\Sigma}_{zz})^2, \tag{8}$$

where $tr(\cdot)$ denotes the trace operator.

Given the results of Theorem 1.3, the potential consequences of serial correlation and conditional heteroscedasticity begin to emerge more clearly. First, consider the traditional scenario where $\tilde{\eta}_{t+k}$ is serially uncorrelated and $\tilde{\varepsilon}_{t+k}$ is conditionally homoscedastic. Under the null, the $\boldsymbol{V}$ matrix in Equation (8) is given by $\sigma_k^2 \Sigma_{zz}^{-1}$, so the mean and variance of Q are equal to m and $2m$, the values for a chi-square distribution with m degrees of freedom. Now let the disturbances display serial correlation and/or conditional heteroscedasticity, and notice how the analysis changes. One difference, of course, is that the matrix $\boldsymbol{V}$ becomes more complex and the results grow less analytically tractable. More importantly, though, the mean and variance of Q take on values that may be far removed from those of a chi-square distribution. This shift in the mean and variance of the limiting distribution of TR_k^2 suggests a possible explanation for reports that long-horizon returns are highly predictable.

Studies that examine long-horizon predictability typically use instrumental variables that are highly persistent. The combination of highly persistent instruments and overlapping returns induces strong serial correlation in the least-squares disturbance vector. As a result, the OLS standard errors for the model understate the variance of the least-squares estimator of the slope coefficients. Although researchers have long recognized that the OLS t-ratios are unreliable in long-horizon regressions, many fail to make the connection between inflated t-ratios and the sample R^2. The easiest way to illustrate the relation is to consider a long-horizon model with a single regressor. In this case, the sample R^2 can be written as $t_{ols}^2/(1 + t_{ols}^2)$, where t_{ols} denotes the OLS t-ratio for a test of the null hypothesis that the least-squares estimator of the slope coefficient is equal to zero. It stands to reason, therefore, that if the OLS t-ratio is inflated, then the sample R^2 will also be misleading.

The intuition for the single-regressor case carries over to the multiple regression setting as well. If the OLS t-ratios are misleading under the null, then the researcher has a clear signal that the traditional interpretation of the sample R^2 is no longer valid. One potential solution to this problem is to use the limiting distribution given in Theorem 1.3 to test whether the sample R^2 is significantly different from zero. Unfortunately, this strategy is not really practical given the complex nature of the distribution in question. There is, however, a related approach that can be implemented quite easily. First, use standard, large-sample, chi-square tests to evaluate whether the null hypothesis of no predictability is credible. In the event that such tests indicate a rejection of the null, then the limiting distribution of the sample R^2 under the alternative hypothesis can be used to draw inferences about the size of the predictable component of returns. The limiting distribution of the sample R^2 under the alternative hypothesis is discussed in Section 1.4.

1.4 Measuring predictability under the alternative

Even if a researcher can reliably reject the null hypothesis that returns are unpredictable, the potential effects of serial correlation and conditional heteroscedasticity remain an important consideration when using the sample R^2 to measure predictability. Let H_A, the alternative hypothesis, be that asset returns are to some extent predictable over time. Any inference concerning the economic significance of the sample R^2 from a predictive regression should be drawn with due consideration for the limiting distribution of this criterion under H_A. The limiting distribution for general situations is provided by Theorem 1.4.

Theorem 1.4. *Let $\{\tilde{r}_t, \tilde{z}_t\}$ be a stationary and ergodic process. Further assume that the regularity conditions given by Hansen (1982) are satisfied. In addition, allow $\tilde{\eta}_{t+k}$ and $\tilde{\varepsilon}_{t+k}$ exhibit the type of serial correlation and conditional heteroscedasticity that is consistent with data generated by a stationary and ergodic process. Define $\tilde{\xi}_{t+k}$, the disturbance term associated with estimating the population R^2, as*

$$\tilde{\xi}_{t+k} \equiv (1 - \rho_k^2)\left(\sum_{i=1}^{k} \tilde{r}_{t+i} - \mu_k\right)^2 - \tilde{\varepsilon}_{t+k}^2, \tag{9}$$

where μ_k is the expected value of the k-period return, and ρ_k^2 denotes the population value of the k-period R^2. Then, under the alternative hypothesis H_A, the limiting distribution of the sample R^2 for the model shown in Equation (1) is

$$\sqrt{T}(R_k^2 - \rho_k^2) \xrightarrow{d} N(0, \sigma_R^2), \tag{10}$$

with σ_R^2 given by

$$\sigma_R^2 \equiv \sum_{j=-\infty}^{\infty} E\left[\frac{\tilde{\xi}_{t+k}\tilde{\xi}_{t+k-j}}{\sigma_k^4}\right], \tag{11}$$

where $\tilde{\xi}_{t+k-j} = (1 - \rho_k^2)(\sum_{i=1}^{k} \tilde{r}_{t+i-j} - \mu_k)^2 - \tilde{\varepsilon}_{t+k-j}^2$.

Theorem 1.4 indicates that under H_A the sample R^2 is asymptotically distributed as a normal random variable. The approach used to derive this result is straightforward. Under the alternative hypothesis, the population value of R^2 lies somewhere between zero and one. Thus, we can easily estimate this population value via the generalized method of moments (GMM), and it follows immediately from the associated distributional theory that the limiting distribution of $\sqrt{T}(R_k^2 - \rho_k^2)$ is normal.

Even though the limiting distribution of the sample R^2 under H_A is normal, the fact that the R^2 criterion is strictly nonnegative suggests that the rate of approach to normality is likely to be relatively slow. It should be possible to improve the approximation, however, by exploiting well-known small sample results. Consider, for example, the scenario where the sequence $\{\tilde{r}_{t+1}, \tilde{z}_t\}$ represents an *i.i.d.* sample from a multivariate normal distribution. Anderson (1984) shows that in this case

$$\sqrt{T}(R_1^2 - \rho_1^2) \overset{d}{\to} N(0, 4\rho_1^2(1 - \rho_1^2)^2). \tag{12}$$

The important thing to note about the distribution in Equation (12) is that the variance is a direct function of the population value of R^2. As a result, the t-ratio for the sample R^2 tends to exhibit substantial departures from normality in small samples. An easy way to improve the small sample performance under these circumstances is to perform a variance-stabilizing transformation.

Suppose we let $f(R_1^2)$ denote the function $\frac{1}{2}ln((1 + R_1)/(1 - R_1))$, where R_1 is the sample multiple-correlation coefficient for a one-period return horizon. It follows by application of the delta method that

$$f(R_1^2) \overset{a}{\sim} N(f(\rho_1^2),\ 1/T). \tag{13}$$

Thus, the transformation yields an asymptotic distribution whose variance is independent of the population value of R^2. Because of this property, there is good reason to suspect that the t-ratio for $f(R_1^2)$ will perform better in small samples than will the t-ratio associated with the original distribution.

Now take the more general scenario where the data can exhibit serial correlation, conditional heteroscedasticity, and nonnormalities. Although the above transformation may no longer achieve complete variance stabilization, it still has the potential to improve inferences about the magnitude of the population R^2. The basic idea is as follows. First compute the quantity $f(R_k^2) = \frac{1}{2}ln((1 + R_k)/(1 - R_k))$. Then construct the standard error of $f(R_k^2)$ using the delta method. The sample multiple correlation coefficient can be written as

$$R_k = \tanh(f(R_k^2)), \tag{14}$$

where $\tanh(\cdot)$ denotes the hyperbolic tangent. Because the hyperbolic tangent is a monotonic transformation, it follows that an approximate $(1 - c)$% confidence interval for the sample multiple correlation coefficient is

$$\tanh(f(\cdot) - \Phi(c/2)\hat{\sigma}_f) \leq \rho_k \leq \tanh(f(\cdot) + \Phi(c/2)\hat{\sigma}_f), \qquad (15)$$

where $\Phi(\cdot)$ and $\hat{\sigma}_f$ denote the cumulative distribution function of a standard normal random variable and the standard error of $f(\cdot)$, respectively.[4]

2. Data and Econometric Methodology

The data used for the empirical analysis mirrors that of Fama and French (1989). This choice of data is motivated by a simple consideration. The Fama and French article is one of the most widely cited studies of long-horizon predictability, so it makes sense to relate the current analysis directly to their work. The dataset is comprised of monthly holding period returns on portfolios of common stock and bonds. The data begin in January 1927 and end in December 1987 (732 observations). All portfolio returns are calculated in excess of the one-month return for the Treasury bill that is closest to 30 days to maturity. Excess returns for horizons exceeding one month are computed by cumulating monthly excess returns.

2.1 The portfolios

The long-term predictability of common stock returns is evaluated using two different market indices: the equal-weighted and value-weighted portfolios constructed from all the firms listed on the New York Stock Exchange (NYSE). Return data for the two indices are provided by the Center for Research in Security Prices (CRSP) at the University of Chicago. The value-weighted index is skewed toward stocks that have a high market capitalization, while the equal-weighted index gives equal representation to all firms. Fama and French (1989) argue that these two portfolios provide a convenient way to study how firm size affects the ability to predict returns.

The bond portfolio returns are drawn from a database maintained by Ibbotson Associates. This database contains monthly holding period returns on a number of different portfolios of corporate bonds. The five portfolios used in the empirical analysis are constructed from bonds rated Aaa, Aa, A, Baa, and below Baa (LG, low-grade) by Moody's. Bond portfolio returns are calculated by taking a price weighted average of the individual returns for the constituent bonds. The average time to maturity of the bonds in each portfolio is typically greater than 10 years.

---

[4] See Chapter 4 of Anderson (1984) for a discussion of this method of drawing inferences within the context of normal correlation theory.

2.2 The Instruments

The instrumental variables used in the long-horizon regressions are the same as those used by Fama and French (1989), but the data used to construct the instruments are drawn from different sources.[5] The first instrument is the dividend yield on the S&P composite stock index (*DIV*). Data on this instrument are obtained from Standard and Poor's *Current Statistics.* The second instrument is a term spread (*TERM*). It is calculated by subtracting the annualized yield on a one-month Treasury bill from the annualized yield on Moody's Aaa bond portfolio. The final instrument is a default risk spread (*DEF*). It is equal to the annualized yield on Moody's Baa bond portfolio minus the annualized yield on Moody's Aaa bond portfolio. The data used to construct both the term spread and the default risk spread are drawn from Moody's *Industrial Manual.* All three instruments are constructed using monthly observations.

Table 1 provides descriptive statistics for the excess returns and instrumental variables. The statistics for the monthly observations are familiar from other studies that use similar data. In general, the monthly stock and bond returns exhibit only a modest degree of serial correlation. The instrumental variables, on the other hand, are highly autocorrelated. Note that the first-order autocorrelation coefficient for both *DIV* and *DEF* exceeds 0.95. It appears, however, that the serial correlation of these two variables decays toward zero at a rate that is consistent with the view that each of the time series is stationary. The results for the annual data are similar.

2.3 The long-horizon regressions

The key claim of Fama and French (1989) is that three instruments—a term spread, a default spread, and the dividend yield on the NYSE index—explain a large fraction of the variation in long-horizon stock and bond returns. This claim stems from the fact that when the portfolio returns are regressed on various combinations of the three instruments, the sample R^2 increases from around 3% for monthly returns to well over 25% for four-year returns. Therefore, the first step is to replicate these regressions by estimating a linear model of the form

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \tilde{z}_t'\beta_k + \tilde{\varepsilon}_{t+k} \tag{16}$$

[5] Fama and French (1989) use the dividend yield on the NYSE index instead of the dividend yield on the S&P composite index, and their term spread and default spread are constructed from yield data maintained by Ibbotson Associates. The advantage of using the Moody's yield data is twofold. First, it is much more widely available than the Ibbotson data and has therefore been used in a number of well-known studies of predictability. Second, its use provides a good check on the general robustness of the Fama and French (1989) results.

Table 1
Summary statistics for the excess returns and instrumental variables

	Mean	SD	Autocorrelations 1	2	3	4	5	6	7	8
Panel A: Monthly data										
VW	0.47	5.77	0.11	−0.01	−0.11	0.03	0.10	−0.01	0.01	0.04
EW	0.75	7.73	0.16	0.02	−0.10	−0.04	0.04	−0.02	0.02	0.01
Aaa	0.06	1.59	0.19	0.00	−0.03	0.00	0.12	0.04	−0.04	0.01
Aa	0.05	1.56	0.24	0.02	0.00	0.00	0.16	0.03	−0.08	0.00
A	0.07	1.96	0.24	0.02	−0.13	−0.01	0.16	0.06	−0.02	0.01
Baa	0.12	2.17	0.22	0.01	−0.14	−0.02	0.13	0.08	−0.03	0.01
LG	0.19	3.22	0.19	0.03	−0.13	−0.08	0.05	0.07	0.08	0.03
DIV	4.46	1.26	0.97	0.93	0.89	0.86	0.83	0.80	0.77	0.73
TERM	2.00	1.36	0.84	0.76	0.73	0.67	0.64	0.64	0.61	0.61
DEF	1.21	0.80	0.97	0.94	0.90	0.89	0.88	0.87	0.86	0.84
Panel B: Annual data										
VW	5.66	20.77	0.10	−0.19	−0.05	−0.12	−0.02	−0.03	0.12	0.06
EW	8.96	29.28	0.14	−0.18	−0.10	−0.18	−0.13	−0.14	0.07	0.04
Aaa	0.73	6.64	0.21	0.05	−0.06	−0.11	−0.17	0.06	−0.03	−0.05
Aa	0.65	6.77	0.20	−0.05	−0.14	−0.16	−0.13	0.03	−0.02	−0.11
A	0.86	8.32	0.25	−0.15	−0.24	−0.14	−0.03	0.11	−0.05	−0.11
Baa	1.43	8.59	0.24	−0.13	−0.24	−0.14	−0.02	0.13	−0.02	−0.07
LG	2.23	12.27	0.32	−0.03	−0.21	−0.21	−0.05	0.11	0.04	0.09
DIV	4.60	1.35	0.57	0.29	0.28	0.26	0.27	0.35	0.30	0.29
TERM	1.90	1.45	0.48	0.20	0.03	0.04	0.22	0.26	0.32	0.12
DEF	1.27	0.87	0.79	0.55	0.37	0.29	0.29	0.36	0.33	0.28

The return data cover the 61-year interval from January 1927–December 1987 (732 monthly observations). Monthly excess returns are defined as the difference between the continuously compounded one-month return on the stock or bond portfolio and the continuously compounded return on a one-month Treasury bill. Annual excess returns are obtained by cumulating monthly excess returns. The stock portfolios are the value-weighted and equal-weighted NYSE indices. Ibbotson forms the bond portfolios by sorting the bonds according to their Moody's ratings: Aaa, Aa, A, Baa, and below Baa (LG, low grade). There are three instrumental variables: the dividend yield on S&P composite index (DIV), the yield on Moody's Aaa bond portfolio less the one-month Treasury bill rate (TERM), and the yield on Moody's Baa bond portfolio less the yield on Moody's Aaa bond portfolio (DEF). Both TERM and DEF are constructed using annualized yields. The observations on the instruments are lagged by one period (i.e., one month in panel A and one year in panel B).

for return horizons of one month, one quarter, and one to four years. Monthly excess returns are defined as the difference between the continuously compounded one-month return on the stock or bond portfolio in question and the continuously compounded return on a one-month Treasury bill. Excess returns for longer return horizons are obtained by cumulating monthly excess returns. The quarterly and one-year returns are nonoverlapping. The two-, three-, and four-year returns are overlapping annual observations.

The methodology used to estimate the regressions is designed to be representative of what is typically seen in the long-horizon literature. Most long-horizon studies use OLS to estimate the models and then rely on some type of correction to obtain consistent standard errors. This approach can easily be nested within a GMM framework.

Consider, for example, an econometric specification based on the disturbance vector,

$$h(\tilde{x}_t, \theta) = \begin{pmatrix} \sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t' \beta_k \\ (\sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t' \beta_k) \tilde{z}_t \end{pmatrix}, \tag{17}$$

where $\tilde{x}_t \equiv [\sum_{i=1}^{k} \tilde{r}_{t+i},\ \tilde{z}_t']'$ and $\theta \equiv [\alpha_k, \beta_k']'$. This disturbance vector contains the OLS normal equations. Because the econometric system in Equation (17) is exactly identified, the GMM estimator of θ is the value that sets $1/T \sum_{t=1}^{T} h(\cdot)$ equal to zero. It is easy to see that this value is obtained by replacing each element of the vector θ with its sample analog. The sample analog of the vector β_k is the least-squares estimator $\hat{\beta}_k$, so the estimates of the slope coefficients obtained using the GMM procedure are identical to those that result from fitting a multiple regression model to the data.

2.4 Tests for predictability

Tests of the null hypothesis that returns are unpredictable are constructed based on the limiting distribution of the GMM estimator.[6] This distribution takes the form

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, (D'S^{-1}D)^{-1}), \tag{18}$$

where

$$D = E\left[\frac{\partial h(\tilde{x}_t, \theta)}{\partial \theta'}\right] \quad \text{and} \quad S = \sum_{j=-\infty}^{\infty} E[h(\tilde{x}_t, \theta) h(\tilde{x}_{t-j}, \theta)']. \tag{19}$$

If θ is partitioned as $[\theta_1'\ \theta_2']'$ where $\theta_2 = \beta_k$ and the matrix $\Omega \equiv (D'S^{-1}D)^{-1}$ is partitioned accordingly, then the Wald statistic,

$$W_D = T(\hat{\theta}_2' \Omega_{22}^{-1} \hat{\theta}_2) \tag{20}$$

converges to a chi-square random variable with m degrees of freedom under the null hypothesis that returns are unpredictable. It is shown in the Appendix that Ω_{22} is the V matrix of Theorem 1.1. Although V is generally unknown, it can be replaced by a consistent estimator without affecting the limiting distribution of the test statistic.

Computing the GMM estimator of the slope coefficients is straightforward, but the approach used to construct the variance-covariance

[6] In the discussion that follows it is assumed that $\tilde{x}_t$ is generated by a stationary, ergodic process and that the second moment matrix of $h(\cdot)$ exists and is finite. This assumption, along with the regularity conditions given by Hansen (1982), is sufficient for the asymptotic distribution theory of GMM to hold.

matrix of this estimator warrants further discussion. The monthly, quarterly, and annual regressions employ nonoverlapping returns, so it quite common to use

$$\hat{V} = \hat{\Sigma}_{zz}^{-1} \left[\frac{1}{T} \sum_{t=1}^{T} \hat{\eta}_{t+k} \hat{\eta}'_{t+k} \right] \hat{\Sigma}_{zz}^{-1} \tag{21}$$

for these models.[7] For the two-, three-, and four-year regressions, the disturbance vector $\tilde{\eta}_{t+k}$ is serially correlated. Thus, it is not immediately obvious how to go about constructing a consistent estimate of V. One possibility, which is discussed in more detail below, is to use a weighting scheme that yields a consistent, positive semidefinite estimator. An alternative approach, which is adopted by Fama and French (1989), is to rely on a truncated estimator. The subsequent analysis makes use of both approaches.

To implement the Fama and French (1989) approach we assume that the nonoverlapping errors for the regression are serially uncorrelated. Under these circumstances, the V matrix of Theorem 1.1 takes the form:

$$V = \Sigma_{zz}^{-1} \left[\sum_{j=-k+1}^{k-1} E[\tilde{\eta}_{t+k} \tilde{\eta}'_{t+k-j}] \right] \Sigma_{zz}^{-1}. \tag{22}$$

As a consequence, we can follow Hansen (1982) and compute a consistent estimate of V from the sample autocovariances of $\tilde{\eta}_{t+k}$ and the sample covariance matrix of the instruments. The main drawback of using this truncated estimator is that it is not guaranteed to be positive semidefinite.

A heteroscedasticity and autocorrelation consistent estimator of V that is guaranteed to be positive semidefinite can be obtained by employing an appropriate weighting scheme. The estimator used in empirical analysis is constructed using the quadratic spectral kernel and automatic bandwidth selection procedure of Andrews (1991).[8] First, the optimal bandwidth is estimated by fitting a first-order autoregressive model to each element of the sample disturbance vector $\hat{\eta}_{t+k}$. This bandwidth controls the rate at which the weights decline

[7] This estimator for V is identical to the heteroscedasticity-consistent covariance matrix estimator of White (1980).

[8] Andrews (1991) shows that quadratic spectral weights are optimal; that is, they minimize the asymptotic truncated mean-squared error of the estimator.

as the lag length increases. Then $\hat{V}$ is computed as

$$\hat{V} = \hat{\Sigma}_{zz}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_{t+k}\hat{\eta}'_{t+k} + \sum_{j=1}^{T-1} w_j\left[\frac{1}{T}\sum_{t=j+1}^{T}(\hat{\eta}_{t+k}\hat{\eta}'_{t+k-j} + \hat{\eta}_{t+k-j}\hat{\eta}'_{t+k})\right]\right]\hat{\Sigma}_{zz}^{-1}, \quad (23)$$

where w_j is the weight at lag j given by the quadratic spectral kernel.

2.5 Standard errors for the sample R^2

Theorem 1.4 provides a way to compute consistent standard errors for the sample R^2. Recall that, under the alternative hypothesis where returns are to some extent predictable over time, the variance of the sample R^2 is a function of the autocovariance structure of $\tilde{\xi}_{t+k}$. In particular, it is given by

$$\sigma_R^2 = \sum_{j=-\infty}^{\infty} E\left[\frac{\tilde{\xi}_{t+k}\tilde{\xi}_{t+k-j}}{\sigma_k^4}\right]. \quad (24)$$

Since the variable $\tilde{\xi}_{t+k}$ will in general be serially correlated, the standard errors for the sample R^2 are computed by applying the automatic bandwidth selection procedure and quadratic spectral kernel of Andrews (1991) to the sample analog of $\tilde{\xi}_{t+k}$.

3. Empirical Results

The empirical section of this article attempts to answer two basic questions. First, are the findings of Fama and French (1989) robust to minor changes in the data used to construct the instrumental variables? Second, and more importantly, can their interpretation of the long-horizon results be justified given the distributional properties of the test statistics and sample R^2? The analysis begins with an overview of the evidence from the long-horizon regressions.

3.1 An initial look at the regression evidence

Tables 2 and 3 report the results of GMM estimation of the multiple regression models. The estimates in Table 2 are for the dividend yield and term spread. Those in Table 3 are for the default spread and term spread. Panel A of each table contains the estimated slopes and their associated t-ratios. As mentioned earlier, these estimates are identical to the ones that would have been obtained by using OLS to fit the models to the data. Panel B of each table presents Wald tests of the null hypothesis that the slopes for the model are equal to zero and gives the sample R^2 for each regression. The t-ratios and Wald tests are corrected for autocorrelation and conditional heteroscedasticity using the truncated estimator of Section 2.4.[9]

Upon initial inspection, the results shown in Tables 2 and 3 would seem to bolster claims that long-horizon stock and bond returns are highly predictable. Almost all of the point estimates of slope coefficients are positive, and many of them are more than two standard errors away from zero. The dividend yield—a stock market variable—seems to have the ability to forecast bond returns. The bond market variables—a term spread and default spread—seem to have the ability to forecast stock returns. Overall, the point estimates of slope coefficients appear to document a clear pattern of time-series variation in expected returns that is common across the stock and bond markets.

The Wald tests and sample R^2 in panel B of Tables 2 and 3 also suggest strong predictability at long horizons. Most of the test statistics for the one-, two-, three-, and four-year returns appear to be highly significant in light of their limiting distribution under the null hypothesis that returns are unpredictable. Moreover, the pattern in the sample R^2 documented by Fama and French (1989) is clearly evident. The sample R^2 increases from less than 3% for monthly returns to well over 25% for many of the four-year returns. In short, the regression results appear to support the conclusion that predictability increases with the length of the return horizon. But a more thorough examination of the regression evidence raises serious questions about whether such a conclusion is actually justified.

3.2 The regression evidence revisited

The fact that the sample R^2 increases with the length of the return horizon does not, in and of itself, signal that long-horizon returns are more predictable than those at short horizons. Indeed, the distributional theory developed earlier suggests that we would *expect* to see an increase in the sample R^2 at long horizons, even if the long-horizon returns are unpredictable. This can easily be illustrated by considering the case where the assumptions of classical regression analysis are satisfied. Under such circumstances, the limiting distribution of the sample R^2 is

$$TR_k^2 \xrightarrow{d} \chi_m^2, \tag{25}$$

where χ_m^2 denotes a chi-square distribution with m degrees of freedom. If, as in the case of monthly returns, there are two instruments

[9] In the regressions where the returns are nonoverlapping, the lag truncation parameter is set equal to zero to obtain White's (1980) estimator of the variance-covariance matrix.

Table 2
Results of regressing excess returns on a dividend yield (DIV) and term spread (TERM)

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DIV_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$$

Panel A: Slopes and t-ratios

Portfolios	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	Slopes for DIV							t-ratios (H_0: $\beta_{1k} = 0$)						
Monthly	0.30	0.58	0.06	0.04	0.00	0.02	0.15	1.15	1.74	1.28	0.59	−0.05	0.17	0.93
Quarterly	1.23	2.27	0.21	0.16	0.10	0.21	0.63	1.34	1.87	1.13	0.80	0.38	0.54	1.16
Annual	3.05	5.73	0.53	0.11	−0.22	−0.05	0.72	1.75	2.64	1.31	0.27	−0.33	−0.07	0.65
Two-Year	8.71	14.82	1.17	0.48	1.18	1.32	4.17	3.56	3.95	1.40	0.54	1.15	1.02	2.05
Three-Year	11.50	18.54	1.94	1.35	3.04	3.34	7.90	3.70	3.57	1.97	1.44	3.26	2.81	3.37
Four-Year	15.00	21.99	2.74	2.23	4.29	4.90	10.65	4.22	3.21	2.50	1.98	3.10	4.09	5.26
	Slopes for TERM							t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	0.15	0.26	0.19	0.19	0.19	0.18	0.21	0.79	1.05	2.58	2.69	2.48	2.14	1.86
Quarterly	0.64	0.96	0.57	0.53	0.61	0.51	0.57	1.12	1.17	2.21	2.01	2.33	1.87	1.59
Annual	1.50	1.93	2.45	2.29	2.54	2.17	2.50	0.84	0.78	5.19	4.17	3.85	3.21	2.58
Two-Year	0.66	0.21	3.10	2.87	3.29	2.97	3.58	0.31	0.06	2.84	2.47	2.25	2.22	1.89
Three-Year	0.31	−0.90	3.26	2.81	3.21	2.78	3.78	0.11	−0.17	2.96	2.25	1.88	1.72	1.82
Four-Year	1.03	1.05	3.31	2.82	2.89	3.01	4.28	0.28	0.20	2.59	1.85	1.38	1.58	1.77

Table 2
(continued)

Portfolios	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DIV_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$													
Panel B: Tests and sample R^2														
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							*p*-values for Wald tests						
Monthly	1.50	3.22	12.59	8.39	6.37	4.73	3.58	0.47	0.20	0.00	0.02	0.04	0.09	0.17
Quarterly	2.09	3.61	9.63	6.13	5.45	3.52	2.91	0.35	0.16	0.01	0.05	0.07	0.17	0.23
Annual	3.50	7.48	27.74	17.54	15.61	10.39	6.80	0.17	0.02	0.00	0.00	0.00	0.01	0.03
Two-Year	14.70	15.85	15.96	9.39	8.84	7.55	8.99	0.00	0.00	0.00	0.01	0.01	0.02	0.01
Three-Year	29.11	12.84	12.32	6.88	15.47	18.90	20.25	0.00	0.00	0.00	0.03	0.00	0.00	0.00
Four-Year	18.70	10.78	9.03	5.37	11.31	19.43	28.96	0.00	0.00	0.01	0.07	0.00	0.00	0.00
	Sample R^2 for the regression (%)													
Monthly	0.6	1.2	2.9	2.8	1.8	1.2	1.2							
Quarterly	2.8	4.3	7.3	5.7	4.8	3.2	3.3							
Annual	5.5	8.4	31.1	24.4	19.4	13.4	9.9							
Two-Year	14.8	20.4	23.0	16.8	15.7	12.9	16.4							
Three-Year	18.9	22.7	18.7	12.9	16.7	14.8	24.5							
Four-Year	26.3	26.8	17.5	13.3	18.5	20.9	32.5							

The regressions for monthly, quarterly, and annual returns use nonoverlapping observations. Those for two-, three-, and four-year returns use overlapping annual observations. There are 732 monthly returns, 244 quarterly returns, 61 annual returns, 60 two-year returns, 59 three-year returns, and 58 four-year returns. The *t*-ratios and Wald tests are corrected for heteroscedasticity and (for multiyear returns) the autocorrelation of overlapping residuals using the methods of Hansen (1982) and White (1980). See note to Table 1 for definition of portfolios.

Table 3
Results of regressing excess returns on a default spread (DEF) and term spread (TERM)

Portfolios	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DEF_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$													
Panel A: Slopes and t-ratios														
	Slopes for DEF							t-ratios (H_0: $\beta_{1k} = 0$)						
Monthly	0.14	0.97	0.08	0.08	0.08	0.08	0.18	0.21	1.08	0.63	0.47	0.30	0.26	0.44
Quarterly	0.63	3.00	0.25	0.26	0.27	0.39	0.67	0.26	0.93	0.57	0.53	0.36	0.39	0.48
Annual	−0.78	6.10	1.39	1.05	0.81	0.68	0.59	−0.23	1.25	2.16	1.23	0.47	0.47	0.25
Two-Year	3.53	17.86	4.40	3.87	5.23	4.11	5.64	0.74	2.58	2.99	1.90	1.91	1.18	1.39
Three-Year	5.58	23.86	7.37	7.33	9.93	8.77	11.64	1.00	2.76	4.51	3.95	4.68	3.01	3.75
Four-Year	7.92	26.88	9.85	9.70	13.11	12.18	15.37	1.02	2.42	4.68	4.62	7.70	5.60	6.21
	Slopes for TERM							t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	0.13	0.03	0.17	0.17	0.17	0.15	0.17	0.72	0.12	1.79	1.86	2.05	1.76	1.54
Quarterly	0.57	0.33	0.52	0.47	0.54	0.42	0.44	1.04	0.43	1.57	1.45	1.97	1.57	1.33
Annual	2.03	0.89	2.13	2.02	2.30	1.99	2.42	1.08	0.33	3.99	3.53	3.80	3.01	2.54
Two-Year	0.68	−2.95	2.04	1.88	2.02	2.01	2.53	0.36	−0.77	1.73	1.56	1.65	1.72	2.03
Three-Year	0.10	−5.24	1.50	0.99	0.88	0.79	1.53	0.04	−1.19	1.23	0.84	0.64	0.52	0.88
Four-Year	0.55	−3.93	0.87	0.37	−0.28	0.16	1.21	0.15	−0.73	0.90	0.34	−0.22	0.12	0.55

Table 3
(continued)

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DEF_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$$

Panel B: Tests and sample R^2

Portfolios	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							p-values for Wald tests						
Monthly	0.88	1.65	11.56	8.61	6.60	5.12	4.24	0.64	0.44	0.00	0.01	0.04	0.08	0.12
Quarterly	1.80	1.69	8.94	5.69	5.62	3.85	3.12	0.41	0.43	0.01	0.06	0.06	0.15	0.21
Annual	1.18	2.18	34.48	18.83	16.99	10.77	7.70	0.55	0.34	0.00	0.00	0.00	0.00	0.02
Two-Year	1.27	7.04	34.19	13.45	13.04	7.78	8.21	0.53	0.03	0.00	0.00	0.00	0.02	0.02
Three-Year	1.07	7.72	27.84	21.93	51.63	20.84	39.13	0.59	0.02	0.00	0.00	0.00	0.00	0.00
Four-Year	1.46	6.34	22.80	22.95	59.28	31.95	42.63	0.48	0.04	0.00	0.00	0.00	0.00	0.00
	Sample R^2 for the regression (%)													
Monthly	0.2	1.0	2.8	2.9	1.9	1.3	1.0							
Quarterly	1.0	2.9	6.9	5.7	4.9	3.2	2.4							
Annual	1.7	4.2	32.6	25.8	19.9	13.8	9.5							
Two-Year	1.4	10.3	31.7	24.7	23.7	16.7	13.5							
Three-Year	1.9	12.9	33.9	30.1	33.5	24.6	20.0							
Four-Year	3.4	13.9	37.1	35.8	41.7	35.1	25.2							

The regressions for monthly, quarterly, and annual returns use nonoverlapping observations. Those for two-, three-, and four-year returns use overlapping annual observations. There are 732 monthly returns, 244 quarterly returns, 61 annual returns, 60 two-year returns, 59 three-year returns, and 58 four-year returns. The t-ratios and Wald tests are corrected for heteroscedasticity and (for multiyear returns) the autocorrelation of overlapping residuals using the methods of Hansen (1982) and White (1980). See note to Table 1 for definition of portfolios.

and 732 observations, then the expected sample R^2 is 0.3%. When we move to annual returns, however, the number of observations drops to 61. As a consequence, the expected sample R^2 increases to 3.3%, even though the population R^2 is equal to zero.

Most of the annual regressions, of course, yield a sample R^2 that is considerably larger than 3.3%. However, the increase in the expected value of the sample R^2 at long horizons is by no means the whole story. We also have to consider the effect that reducing the number of observations has on the standard deviation of the sample R^2. Like the expected value, the standard deviation rises from 0.3% for monthly returns to 3.3% for annual returns. Moreover, the sample R^2 for an annual regression would have to exceed 9.8% in order for us to reject the null of no predictability at a 5% significance level. Only 9 of the 14 annual regressions yield a sample R^2 that exceeds this cutoff, and we have yet to take the effects of conditional heteroscedasticity and serial correlation into account.

To properly gauge the precision of the long-horizon estimates we need standard errors for the sample R^2 that have been corrected for autocorrelation and conditional heteroscedasticity. These are shown in Table 4. Panel A looks at the dividend yield and term spread regressions. The default and term spread regressions are covered by panel B. Note that the standard errors for the monthly regressions are relatively small, but those for the long-horizon regressions often exceed 10%. This is particularly true for the high-grade bond portfolios, which are the same portfolios that produce the largest values of the sample R^2. It appears that, in general, as the sample R^2 increases, so does its standard error.

Table 4 also reports lower confidence limits for 95% and 98% confidence intervals on the population R^2. These lower confidence limits provide a way to get a better feel for the impact of the observed increase in the standard errors at long horizons. They are constructed using the variance-stabilizing transformation of the limiting distribution of the sample multiple correlation coefficient that was discussed in Section 1.4. A lower confidence limit of zero indicates that it is not possible to reject the hypothesis that the multiple correlation coefficient, and hence the population R^2, for the regression is equal to zero.

On balance, the lower confidence limits shown in Table 4 do not provide a lot of support for the view that the ability to predict stock and bond returns increases with the length of the return horizon. Take the results for the value-weighted market portfolio as a case in point. At the four-year horizon, the dividend yield and term spread regression yields a sample R^2 of 26.3%. Although this value may appear

Table 4
Standard errors and lower confidence limits for the sample R^2 from the multiple regressions

	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
Panel A: DIV and TERM														
	Sample R^2 for the regression							Standard errors for the sample R^2						
Monthly	0.6	1.2	2.9	2.8	1.8	1.2	1.2	0.9	1.2	2.5	2.6	1.7	1.4	1.3
Quarterly	2.8	4.3	7.3	5.7	4.8	3.2	3.3	3.0	3.5	5.7	5.1	4.0	3.4	3.4
Annual	5.5	8.4	31.1	24.4	19.4	13.4	9.9	6.1	5.9	12.7	11.6	12.3	9.5	8.7
Two-Year	14.8	20.4	23.0	16.8	15.7	12.9	16.4	7.3	8.7	11.1	10.7	10.3	9.0	9.0
Three-Year	18.9	22.7	18.7	12.9	16.7	14.8	24.5	8.9	9.5	11.3	10.1	9.8	8.4	9.5
Four-Year	26.3	26.8	17.5	13.3	18.5	20.9	32.5	11.5	13.1	12.4	11.7	11.4	10.4	11.6
	95% lower confidence limit							98% lower confidence limit						
Monthly	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Quarterly	0.0	0.1	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Annual	0.0	0.6	8.8	5.4	1.8	0.8	0.1	0.0	0.2	5.8	3.2	0.5	0.1	0.0
Two-Year	3.4	6.0	5.0	1.6	1.4	0.9	2.8	2.1	4.1	2.9	0.5	0.4	0.2	1.4
Three-Year	4.6	6.7	2.2	0.4	2.3	2.3	8.1	2.9	4.6	0.8	0.0	1.0	1.1	5.9
Four-Year	6.9	5.3	0.9	0.0	2.0	4.4	11.4	4.4	2.9	0.1	0.0	0.7	2.5	8.3

Table 4
(continued)

	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
Panel B: DEF and TERM														
	Sample R^2 for the regression							Standard errors for the sample R^2						
Monthly	0.2	1.0	2.8	2.9	1.9	1.3	1.0	0.5	1.6	2.4	2.5	1.9	1.5	1.3
Quarterly	1.0	2.9	6.9	5.7	4.9	3.2	2.4	1.8	3.7	5.5	5.1	4.3	3.7	3.1
Annual	1.7	4.2	32.6	25.8	19.9	13.8	9.5	2.1	5.5	12.0	11.7	13.8	11.7	8.5
Two-Year	1.4	10.3	31.7	24.7	23.7	16.7	13.5	3.0	8.9	7.8	9.0	13.1	10.7	9.0
Three-Year	1.9	12.9	33.9	30.1	33.5	24.6	20.0	4.1	10.5	10.7	10.0	10.8	10.7	8.7
Four-Year	3.4	13.9	37.1	35.8	41.7	35.1	25.2	6.3	12.7	12.5	12.8	10.5	10.0	10.1
	95% lower confidence limit							98% lower confidence limit						
Monthly	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Quarterly	0.0	0.0	0.2	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Annual	0.0	0.0	10.9	6.3	1.0	0.1	0.0	0.0	0.0	7.8	3.9	0.1	0.0	0.0
Two-Year	0.0	0.1	17.0	9.0	3.4	1.6	1.2	0.0	0.0	14.5	6.7	1.5	0.5	0.3
Three-Year	0.0	0.2	14.0	12.0	13.5	6.6	5.7	0.0	0.0	10.9	9.2	10.4	4.3	3.8
Four-Year	0.0	0.0	13.6	12.2	21.0	16.2	7.9	0.0	0.0	10.0	8.6	17.4	13.1	5.5

The regressions for monthly, quarterly, and annual returns use nonoverlapping observations. Those for two-, three-, and four-year returns use overlapping annual observations. There are 732 monthly returns, 244 quarterly returns, 61 annual returns, 60 two-year returns, 59 three-year returns, and 58 four-year returns. The standard errors are corrected for heteroscedasticity and autocorrelation using the procedures outlined in Andrews (1991). The lower confidence limits are constructed using a variance stabilizing transformation of the limiting distribution of the sample multiple correlation coefficient. See note to Table 1 for definition of portfolios.

large in absolute magnitude, the lower limit for the 95% confidence interval is only 6.5%. Thus, a sample R^2 of 26.3% does not necessarily imply that four-year returns on the market portfolio are highly predictable.

Of course, several of the lower confidence limits in Table 4 are a good deal higher than 6.5%. The A-rated bond portfolio, for example, yields a lower limit of 21% at the four-year horizon. But we have to be careful not to place undue emphasis on the outcome of a single regression. Searching through the table to find the regression that yields the largest lower confidence limit is a form of data snooping. A number of studies, such as those by Foster, Smith, and Whaley (1996) and Lo and MacKinlay (1995), show that even a modest degree of data snooping can severely bias classical procedures for drawing inferences. Basically it boils down to a question of whether the regression evidence, when taken as a whole, supports the conclusion that predictability increases at long horizons. The pattern of standard errors in Table 4 suggests that the increase in the sample R^2 may to a large extent be a small sample effect. If this is the case, however, then there should also be a way to explain the apparent evidence of strong predictability from the t-ratios and Wald tests.

3.3 Size and power considerations

There are a couple of plausible scenarios under which we could observe large t-ratios and Wald statistics for the long-horizon regressions even though long-horizon returns are not highly predictable. The first is where the long-horizon returns are unpredictable, but the t-ratios and Wald tests exhibit poor size. The second is where long-horizon returns are only slightly predictable, but the t-ratios and Wald tests exhibit high power. Recent studies by Richardson and Smith (1991) and Hodrick (1992) demonstrate that size is a definite concern in the long-horizon setting. These studies also suggest that the size of the tests can be improved by explicitly imposing the null when constructing an estimate of the variance-covariance matrix of the OLS estimator of the slope coefficients.

Tables 5 and 6 illustrate the effect that imposing the null has on the t-ratios and Wald tests. The results in panel A are for the truncated estimator. Those in panel B are for the quadratic spectral estimator of Andrews (1991). The procedure used to impose the null when constructing an estimate of the variance-covariance matrix is straightforward. Under the null, the slope coefficients in the regression model are equal to zero, so the disturbance vector $\tilde{\eta}_{t+k}$ is given by $\sum_{i=1}^{k} \tilde{r}_{t+i}(\tilde{z}_t - \boldsymbol{\mu}_z)$. As a result, we can compute the truncated estimator as

$$\hat{V} = \hat{\Sigma}_{zz}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_{t+k}\hat{\eta}'_{t+k} + \sum_{j=1}^{k-1}\left[\frac{1}{T}\sum_{t=j+1}^{T}(\hat{\eta}_{t+k}\hat{\eta}'_{t+k-j} + \hat{\eta}_{t+k-j}\hat{\eta}'_{t+k})\right]\right]\hat{\Sigma}_{zz}^{-1}, \quad (26)$$

and the quadratic spectral estimator as

$$\hat{V} = \hat{\Sigma}_{zz}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_{t+k}\hat{\eta}'_{t+k} + \sum_{j=1}^{T-1}w_j\left[\frac{1}{T}\sum_{t=j+1}^{T}(\hat{\eta}_{t+k}\hat{\eta}'_{t+k-j} + \hat{\eta}_{t+k-j}\hat{\eta}'_{t+k})\right]\right]\hat{\Sigma}_{zz}^{-1}, \quad (27)$$

where $\hat{\eta}_{t+k}$ denotes the sample analog of $\sum_{i=1}^{k} \tilde{r}_{t+i}(\tilde{z}_t - \boldsymbol{\mu}_z)$, and w_j denotes the weight at lag j given by the quadratic spectral kernel using the bandwidth determined by fitting an AR(1) model to each element of $\hat{\eta}_{t+k}$.

It is immediately apparent that the t-ratios in panel A of Tables 5 and 6 are markedly different from those shown in Tables 2 and 3. In Table 2, for instance, the annual regressions for the *Aaa*, *Aa*, and *A* bond portfolios produce t-ratios for the term spread of 5.19, 4.17, and 3.85. The corresponding values in panel A of Table 5 are 2.70, 2.68, and 2.77. This difference is solely a function of whether or not the null is imposed when constructing the variance-covariance matrix. When the null is imposed, few of the long-horizon regressions yield t-ratios that are greater than two in magnitude, and most of those that do involve one-year returns. The Wald statistics that result from imposing the null are also much smaller than those reported in the previous tables. Most of Wald tests in panel A of Tables 5 and 6 do not reject the null of no predictability at conventional significance levels.

One potential concern is that these findings have something to do with the fact that the truncated estimator is not guaranteed to be positive semidefinite. This is not the case, however. Panel B of Tables 5 and 6 show that the t-ratios and Wald statistics for spectral estimator are quite similar to those for the truncated estimator. Thus, the fact that the truncated estimator is not guaranteed to be positive semidefinite does not appear to play a meaningful role in the results. The spectral estimator of the variance-covariance matrix does tend to yield substantially smaller t-ratios and test statistics for annual returns.

Table 5
Regression results for DIV and TERM using covariance matrix estimators that impose the null

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DIV_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$$

Panel A: Truncated estimator

	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	DIV t-ratios (H_0: $\beta_{1k} = 0$)							TERM t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	1.14	1.69	1.26	0.58	−0.05	0.17	0.91	0.78	1.03	2.55	2.68	2.47	2.12	1.84
Quarterly	1.24	1.67	1.00	0.73	0.36	0.51	1.07	1.05	1.08	2.05	1.91	2.20	1.77	1.52
Annual	1.83	2.47	0.91	0.21	−0.33	−0.07	0.07	0.74	0.70	2.70	2.68	2.77	2.53	2.28
Two-Year	2.68	2.57	0.96	0.45	0.98	0.96	1.74	0.30	0.05	1.97	1.87	1.92	1.85	1.68
Three-Year	2.51	2.36	1.31	1.11	2.07	2.09	2.23	0.09	−0.15	1.70	1.56	1.43	1.32	1.29
Four-Year	2.12	1.94	1.65	1.50	1.93	2.11	2.18	0.26	0.19	1.62	1.37	1.12	1.21	1.27
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							p-values for Wald tests						
Monthly	1.44	3.00	11.47	8.07	6.39	4.71	3.47	0.49	0.22	0.00	0.02	0.04	0.09	0.18
Quarterly	1.77	2.85	6.86	4.90	4.82	3.18	2.55	0.41	0.24	0.03	0.09	0.09	0.20	0.28
Annual	3.87	6.56	9.03	8.06	7.70	6.48	5.73	0.14	0.04	0.01	0.02	0.02	0.04	0.06
Two-Year	8.77	6.64	6.05	5.43	5.67	6.43	6.58	0.01	0.04	0.05	0.07	0.06	0.04	0.04
Three-Year	6.71	6.52	3.68	2.95	4.40	5.31	5.26	0.03	0.04	0.16	0.23	0.11	0.07	0.07
Four-Year	4.62	4.09	2.99	2.38	3.74	4.44	4.74	0.10	0.13	0.22	0.30	0.15	0.11	0.09

Table 5
(continued)

	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DIV_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$													
Panel B: Spectral estimator														
	DIV t-ratios (H_0: $\beta_{1k} = 0$)							TERM t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	1.17	1.67	1.16	0.58	−0.05	0.18	0.98	0.84	1.05	2.00	2.06	2.12	1.94	1.76
Quarterly	1.52	1.92	1.08	0.79	0.44	0.63	1.32	1.20	1.33	2.06	1.95	2.29	2.00	1.84
Annual	1.67	2.31	0.96	0.22	−0.28	−0.06	0.57	0.99	0.84	1.64	1.66	1.96	1.92	1.88
Two-Year	2.87	2.81	1.03	0.47	1.04	1.00	1.80	0.27	0.05	1.99	1.93	2.05	1.90	1.73
Three-Year	2.74	2.44	1.32	1.07	1.88	2.01	2.29	0.09	−0.15	1.72	1.59	1.52	1.43	1.33
Four-Year	2.65	2.26	1.60	1.46	2.08	2.31	2.38	0.26	0.17	1.70	1.46	1.20	1.40	1.32
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							p-values for Wald tests						
Monthly	1.59	3.03	6.53	4.85	4.54	3.88	3.44	0.45	0.22	0.04	0.09	0.10	0.14	0.18
Quarterly	2.74	3.92	6.57	4.80	5.25	4.00	3.83	0.25	0.14	0.04	0.09	0.07	0.14	0.15
Annual	3.66	5.73	5.29	4.11	3.93	3.76	4.13	0.16	0.06	0.07	0.13	0.14	0.15	0.13
Two-Year	9.64	7.89	5.83	5.59	6.40	6.87	6.48	0.01	0.02	0.05	0.06	0.04	0.03	0.04
Three-Year	8.26	6.30	3.92	3.29	4.67	6.56	6.05	0.02	0.04	0.14	0.19	0.10	0.04	0.05
Four-Year	7.07	5.89	3.45	2.70	4.33	5.60	5.69	0.03	0.05	0.18	0.26	0.11	0.06	0.06

The regressions for monthly, quarterly, and annual returns use nonoverlapping observations. Those for two-, three-, and four-year returns use overlapping annual observations. There are 732 monthly returns, 244 quarterly returns, 61 annual returns, 60 two-year returns, 59 three-year returns, and 58 four-year returns. The t-ratios and Wald tests are computed by imposing the null that the slopes are equal to zero when constructing the estimate of the variance-covariance matrix. They are corrected for heteroscedasticity and (for multiyear returns) the autocorrelation of overlapping residuals using: (i) the methods of Hansen (1982) and White (1980) in panel A; (ii) the procedures outlined in Andrews (1991) in panel B. See note to Table 1 for definition of portfolios.

Table 6
Regression results for DEF and TERM using covariance matrix estimators that impose the null

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DEF_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$$

Panel A: Truncated estimator

	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	DEF t-ratios (H_0: $\beta_{1k} = 0$)							TERM t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	0.21	1.04	0.62	0.47	0.30	0.25	0.43	0.72	0.12	1.78	1.88	2.08	1.76	1.55
Quarterly	0.25	0.84	0.49	0.48	0.34	0.37	0.45	1.00	0.43	1.46	1.39	1.83	1.47	1.30
Annual	−0.22	1.19	0.96	0.89	0.50	0.53	0.28	0.92	0.33	1.97	2.08	2.54	2.24	2.36
Two-Year	0.61	1.39	1.58	1.65	1.46	1.39	1.23	0.33	−1.02	1.30	1.27	1.45	1.45	1.80
Three-Year	0.74	1.35	1.71	1.81	1.77	2.03	1.60	0.04	−1.68	0.98	0.78	0.71	0.56	1.13
Four-Year	0.69	1.16	1.79	1.79	1.79	1.97	1.56	0.15	−0.75	0.83	0.31	−0.17	0.10	0.51
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							p-values for Wald tests						
Monthly	0.87	1.51	10.61	8.22	6.57	5.07	4.24	0.65	0.47	0.00	0.02	0.04	0.08	0.12
Quarterly	1.67	1.46	6.80	4.78	4.92	3.46	2.93	0.43	0.48	0.03	0.09	0.09	0.18	0.23
Annual	0.87	1.78	11.34	10.10	7.63	6.56	5.94	0.65	0.41	0.00	0.01	0.02	0.04	0.05
Two-Year	0.71	2.53	5.73	5.41	4.22	4.29	3.80	0.70	0.28	0.06	0.07	0.12	0.12	0.15
Three-Year	0.56	6.56	3.70	3.45	3.13	4.16	2.55	0.76	0.04	0.16	0.18	0.21	0.13	0.28
Four-Year	0.62	1.41	3.21	3.28	3.33	4.05	2.61	0.73	0.50	0.20	0.19	0.19	0.13	0.27

Table 6
(continued)

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} DEF_t + \beta_{2k} TERM_t + \tilde{\varepsilon}_{t+k}$$

Panel B: Spectral estimator	VW	EW	Aaa	Aa	A	Baa	LG	VW	EW	Aaa	Aa	A	Baa	LG
	DEF t-ratios (H_0: $\beta_{1k} = 0$)							TERM t-ratios (H_0: $\beta_{2k} = 0$)						
Monthly	0.22	1.04	0.54	0.44	0.31	0.28	0.50	0.70	0.12	1.44	1.49	1.68	1.55	1.33
Quarterly	0.33	1.04	0.53	0.53	0.44	0.51	0.62	1.01	0.46	1.52	1.45	1.94	1.59	1.49
Annual	−0.23	1.15	0.80	0.84	0.56	0.48	0.27	1.23	0.45	1.29	1.37	1.82	1.76	2.10
Two-Year	0.66	1.36	1.55	1.69	1.62	1.61	1.41	0.31	−1.00	1.39	1.40	1.58	1.47	1.82
Three-Year	0.66	1.21	1.73	1.83	1.74	2.04	1.51	0.04	−1.51	0.93	0.71	0.59	0.51	0.96
Four-Year	0.71	1.17	1.82	1.80	1.75	1.96	1.51	0.16	−0.86	0.69	0.28	−0.17	0.10	0.59
	Wald tests (H_0: β_{1k} and $\beta_{2k} = 0$)							p-values for Wald tests						
Monthly	0.98	1.70	6.46	5.28	4.49	3.92	3.48	0.61	0.43	0.04	0.07	0.11	0.14	0.18
Quarterly	1.83	2.25	6.66	4.86	5.26	4.04	3.87	0.40	0.33	0.04	0.09	0.07	0.13	0.14
Annual	1.62	1.65	4.38	4.60	3.94	3.73	4.45	0.45	0.44	0.11	0.10	0.14	0.15	0.11
Two-Year	0.62	2.91	4.44	4.27	3.94	4.68	3.83	0.73	0.23	0.11	0.12	0.14	0.10	0.15
Three-Year	0.47	2.67	3.75	3.58	3.06	4.16	2.72	0.79	0.26	0.15	0.17	0.22	0.12	0.26
Four-Year	0.69	1.46	3.42	3.24	3.09	3.88	2.83	0.71	0.48	0.18	0.20	0.21	0.14	0.24

The regressions for monthly, quarterly, and annual returns use nonoverlapping observations. Those for two-, three-, and four-year returns use overlapping annual observations. There are 732 monthly returns, 244 quarterly returns, 61 annual returns, 60 two-year returns, 59 three-year returns, and 58 four-year returns. The t-ratios and Wald tests are computed by imposing the null that the slopes are equal to zero when constructing the estimate of the variance-covariance matrix. They are corrected for heteroscedasticity and (for multiyear returns) the autocorrelation of overlapping residuals using: (i) the methods of Hansen (1982) and White (1980) in panel A; (ii) the procedures outlined in Andrews (1991) in panel B. See note to Table 1 for definition of portfolios.

This is not surprising given that the spectral estimator corrects for serial correlation in $\tilde{\eta}_{t+k}$. As Table 1 shows, the annual returns seem to exhibit a modest degree of autocorrelation. The spectral estimator captures this autocorrelation, and the truncated estimator does not.

After reviewing the evidence from the tests that impose the null, it is not obvious how to interpret the results of the long-horizon regressions. We know that, under the null hypothesis, the tests that impose the null and those that do not have exactly the same limiting distribution. But this does not necessarily imply that the size of the tests is the same in small samples. It could be that the conventional tests reject the null too often when it is true. This would provide support for using the t-ratios and Wald statistics that impose the null. At the same time, however, we have to allow for the possibility that the tests that impose the null have low power in the long-horizon setting. In this case, the conventional t-ratios and Wald statistics might provide more reliable results. A Monte Carlo experiment provides a convenient way to shed additional light on such issues.

4. Monte Carlo Evidence

The results of the empirical analysis suggest that previous research overstates the ability to predict long-horizon stock and bond returns. It is important to recognize, however, that this inference is based on asymptotic distribution theory. This could have a significant impact on the findings because a number of studies have shown that for datasets of the size used in this study, the small sample distribution of long-horizon test statistics can exhibit significant departures from the theoretical limiting distribution [see, for example, Richardson and Stock (1989), Kim, Nelson, and Startz (1991), Richardson and Smith (1991); Goetzmann and Jorion (1993, 1995), and Nelson and Kim (1993)]. This section reports the results of a series of Monte Carlo experiments that are designed to assess whether the small sample behavior of the test statistics and sample R^2 is cause for concern.

4.1 The data generating process

Choosing a data generating process for the simulations involves a trade-off between two opposing considerations. First and foremost, the process should be capable of closely reproducing the observed characteristics of the actual returns. At the same time, it should not be overly complex. The simulations in this section assume that returns are generated by a linear factor model where the factor loading is a linear function of the instruments. This process is motivated by a couple of observations. First, it is consistent with an intertemporal version of the APT of Ross (1976). With a constant price of risk and a factor loading that is linear in $\tilde{z}$, the intertemporal APT implies that there should be a linear relation between expected returns and realizations of the instrumental variables. Second, the process can produce strong predictability at long horizons that stems directly from persistence in short-horizon expected returns. This is precisely the sort of predictability that Fama and French (1989) contend exists in actual returns.

The exact form of the process that is assumed to govern the evolution of the returns and instrumental variables is

$$\begin{aligned} \tilde{r}_t &= m_t + b_t \tilde{f}_t + \psi \tilde{u}_t & \psi &> 0 \\ \tilde{z}_{it} &= \phi \tilde{z}_{i,t-1} + \gamma \tilde{\eta}_{it} & \gamma &> 0,\ -1 < \phi < 1,\ i = 1, 2 \\ m_t &= \lambda b_t & \lambda &> 0 \\ b_t &= \omega + \delta_1 \tilde{z}_{1,t-1} + \delta_2 \tilde{z}_{2,t-1} & \omega,\ \delta_1,\ \delta_2 &\geq 0 \end{aligned} \tag{28}$$

where

$$[\tilde{f}_t,\ \tilde{u}_t,\ \tilde{\eta}_{1t},\ \tilde{\eta}_{2t}]' \sim i.i.d.\ N(\mathbf{0},\ \mathbf{I}).$$

This specification captures the pertinent features of long-horizon regressions in a parsimonious fashion. Returns are given by the sum of a stationary component and a noise component, which is similar to return process implied by the two-component models of stock prices proposed by Summers (1986), Fama and French (1988b), and Porteba and Summers (1988). It differs from such models, however, in one notable respect. Returns generated under the null of no predictability are homoscedastic. But under the alternative where returns are predictable, they exhibit conditional heteroscedasticity. Moreover, for ϕ close to one, the process gives rise to autoregressive conditional heteroscedasticity (ARCH) effects like those that have been widely documented in empirical literature.

To see how Equation (28) gives rise to ARCH effects, note that the conditional variance of returns is given by $b_t^2 + \psi^2$. Setting ϕ close to one causes b_t^2 to be strongly autocorrelated, and the resulting conditional variance process exhibits a high degree of persistence. Note also that the conditional mean and the conditional variance will in general be correlated with one another. The parameter ω controls the degree of correlation. If the other parameters are held constant, then the correlation between the conditional mean and conditional variance increases as ω gets larger. Thus, the process is similar in some respects to an ARCH-in-mean specification.

4.2 Parameter settings

To conduct the simulations it is necessary to select settings for each of the parameters in Equation (28). The settings used to generate the data for the case where returns are unpredictable are as follows: $\psi = 10$, $\lambda = 0.5$, $\omega = 10$, $\delta_1 = 0$, $\delta_2 = 0$, and $\gamma = \sqrt{1-\phi^2}$. These values are consistent with a scenario where the expected excess return on the asset is 5% and the return standard deviation is 14.14%. The impact of serial correlation on the analysis is gauged by using four different settings of the parameter ϕ. These settings are 0.0, 0.3, 0.6, and 0.9. The parameter γ is set equal to $\sqrt{1-\phi^2}$ so that the instrumental variables always have unit variance.

Returns generated using the aforementioned parameter settings are *i.i.d.* normal random variables that are distributed independently of the instrumental variables. The length of the return series used in the simulations is the same as that for the actual long-horizon regressions. It ranges from a high of 61 for the one-period returns to a low of 58 for the four-period returns. The one-period returns are nonoverlapping. Those for two-, three-, and four-periods are overlapping one-period observations. The basic idea behind the simulations is to estimate a linear regression model of the form

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k}\tilde{z}_{1t} + \beta_{2k}\tilde{z}_{2t} + \tilde{\varepsilon}_{t+k} \tag{29}$$

a total of 10,000 times for each of the 16 possible combinations of return horizon and parameter settings. The results of the simulation are then used to infer how the test statistics and sample R^2 will perform in long-horizon regressions where the returns are unpredictable.

The procedure used to generate data for the situation where returns exhibit predictability is more complicated than for the case where they are unpredictable. The strategy is to choose a desired level of predictability, and then find a combination of parameter settings that will deliver it. For the simulations, this is accomplished as follows: (i) set the population R^2 to 5%; (ii) set the parameters ψ, ω, and γ to 10, 10, and $\sqrt{1-\phi^2}$, respectively; (iii) set δ_1 equal to either 5 or 10; (iv) set $\delta_2 = 0$; (v) set ϕ equal to 0.3, 0.6, or 0.9; (vi) solve for the value of λ that delivers a population R^2 of 5%. After generating the data, estimation of the regression model proceeds along the same lines as before.

To solve for the correct value of λ we need a way to compute the population R^2. With δ_2 set equal to zero, the general formula for the population R^2 for a k-period regression is

$$\rho_k^2 \equiv \left(\frac{\operatorname{cov}(\sum_{i=1}^{k} \tilde{r}_{t+i}, \tilde{z}_{1t})^2}{\operatorname{var}(\sum_{i=1}^{k} \tilde{r}_{t+i})\operatorname{var}(\tilde{z}_{1t})} \right). \tag{30}$$

This equation can be expressed as

$$\rho_k^2 = \rho_1^2 \left(\frac{1-\phi^k}{1-\phi}\right)^2 \left(\frac{\operatorname{var}(\sum_{i=1}^{k} \tilde{r}_{t+i})}{\operatorname{var}(\tilde{r}_{t+1})}\right)^{-1}, \tag{31}$$

where ρ_1^2 is the value of the population R^2 for the one-period regression. Thus, the population R^2 for the k-period regression depends on the population R^2 for the one-period model, the autocorrelation coefficient of the first instrument, and the k-period variance ratio. For the data generating process in Equation (28), Equation (31) simplifies to

$$\rho_k^2 = \left(\frac{\left(\frac{1-\phi^k}{1-\phi}\right)^2 \left(\frac{\lambda^2\delta_1^2\gamma^2}{(\psi^2+\omega^2)(1-\phi^2)+\gamma^2\delta_1^2(1+\lambda^2)}\right)}{k + 2\left(\frac{\lambda^2\delta_1^2\gamma^2}{(\psi^2+\omega^2)(1-\phi^2)+\gamma^2\delta_1^2(1+\lambda^2)}\right)\sum_{i=1}^{k-1}(k-i)\phi^i} \right). \tag{32}$$

Equation (32) can be solved to find the value of λ that yields the desired level of predictability.[10] The various combinations of parameter settings outlined in steps (i) through (v) yield values of λ that range from a low of 0.232 to a high of 1.03.

4.3 The evidence on size

Table 7 examines the distributional properties of the t-ratios and sample R^2 for the situation where the population R^2 is equal to zero. Recall that this implies that the simulated returns are conditionally homoscedastic and distributed independently of the instruments. The table reports the mean, standard deviation, and percentage points for three different statistics: the t-ratio computed using the conventional truncated estimator of the variance-covariance matrix, the t-ratio based on the truncated estimator that imposes the null, and the sample R^2 for the regression model. The null hypothesis for the t-ratios is that the slope coefficient for the first instrumental variable is equal to zero.

The results shown in the initial four lines of the table are for regressions that use one-period returns. Because the one-period returns are nonoverlapping and conditionally homoscedastic, the traditional assumptions of linear regression analysis are satisfied. As a consequence, we might expect that the empirical distributions of the t-ratios and sample R^2 would be well approximated by their theoretical lim-

---

[10] See the Appendix for the details of how this equation is derived.

Table 7
Percentage points for the slope t-ratios and sample R^2 in simulations where the population R^2 is zero

Horizon	ϕ	0.01	0.025	0.50	0.975	0.99	0.01	0.025	0.50	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
		Conventional t-ratio (H_0: $\beta_{1k} = 0$)					t-ratio that imposes H_0: $\beta_{1k} = 0$					Sample R^2 for the regression						
1-period	0.0	−2.63	−2.21	−0.02	2.14	2.61	−2.30	−1.98	−0.02	1.94	2.28	3.3	3.2	2.4	7.6	9.7	12.0	14.7
1-period	0.3	−2.62	−2.17	−0.01	2.19	2.65	−2.30	−1.97	−0.01	1.96	2.31	3.3	3.2	2.3	7.7	9.9	12.0	14.4
1-period	0.6	−2.59	−2.16	0.00	2.19	2.61	−2.28	−1.97	0.00	1.97	2.28	3.3	3.3	2.3	7.7	9.9	12.1	14.9
1-period	0.9	−2.60	−2.18	0.01	2.14	2.53	−2.30	−1.99	0.01	1.95	2.24	3.4	3.3	2.4	7.7	9.9	12.2	15.0
2-periods	0.0	−2.82	−2.32	0.00	2.32	2.83	−2.26	−1.97	0.00	1.95	2.23	3.3	3.2	2.4	7.7	10.0	11.9	14.6
2-periods	0.3	−2.90	−2.37	0.01	2.37	2.90	−2.22	−1.95	0.01	1.95	2.22	4.3	4.1	3.0	9.8	12.6	15.3	18.1
2-periods	0.6	−2.91	−2.40	0.01	2.42	2.90	−2.18	−1.91	0.01	1.91	2.19	5.2	4.9	3.7	11.8	15.2	18.3	22.0
2-periods	0.9	−2.99	−2.43	0.02	2.39	2.94	−2.13	−1.89	0.02	1.89	2.16	6.1	5.6	4.5	13.7	17.5	21.2	25.6
3-periods	0.0	−3.04	−2.43	0.02	2.45	3.03	−2.21	−1.92	0.02	1.95	2.22	3.4	3.3	2.4	7.9	10.0	12.3	14.8
3-periods	0.3	−3.06	−2.53	0.01	2.52	3.19	−2.15	−1.92	0.01	1.90	2.17	4.8	4.6	3.5	11.0	14.1	17.2	20.7
3-periods	0.6	−3.24	−2.67	0.00	2.71	3.26	−2.14	−1.91	0.00	1.88	2.12	6.6	6.1	4.8	15.0	18.9	22.8	26.5
3-periods	0.9	−3.47	−2.74	0.00	2.79	3.58	−2.08	−1.85	0.00	1.87	2.10	8.6	7.6	6.4	19.1	24.2	28.8	33.8

Table 7
(continued)

Horizon	ϕ	0.01	0.025	0.50	0.975	0.99	0.01	0.025	0.50	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
		Conventional t-ratio (H_0: $\beta_{1k} = 0$)					t-ratio that imposes H_0: $\beta_{1k} = 0$					Sample R^2 for the regression						
4-periods	0.0	−3.35	−2.61	0.00	2.63	3.42	−2.20	−1.92	0.00	1.93	2.20	3.5	3.5	2.4	8.1	10.4	12.7	15.9
4-periods	0.3	−3.46	−2.75	0.00	2.76	3.48	−2.13	−1.90	0.00	1.89	2.12	5.2	4.9	3.7	12.1	15.1	18.3	22.1
4-periods	0.6	−3.62	−2.86	0.00	2.94	3.76	−2.09	−1.87	0.00	1.87	2.09	7.7	6.9	5.6	17.4	21.8	25.9	30.6
4-periods	0.9	−4.03	−3.07	0.00	3.14	3.96	−2.01	−1.82	0.00	1.81	2.04	10.9	9.4	8.3	24.0	29.6	35.2	40.5
Normal (0,1)		−2.33	−1.96	0.00	1.96	2.33	−2.33	−1.96	0.00	1.96	2.33							

The data generating process for excess returns is a conditional linear factor model of the form

$$\tilde{r}_t = m_t + b_t\tilde{f}_t + \psi\tilde{u}_t,$$

where (i) $m_t = \lambda b_t$; (ii) $b_t = \omega + \delta_1\tilde{z}_{1,t-1} + \delta_2\tilde{z}_{2,t-1}$; (iii) $\tilde{z}_{1t} = \phi\tilde{z}_{1,t-1} + \gamma\tilde{\eta}_{1t}$; (iv) $\tilde{z}_{2t} = \phi\tilde{z}_{2,t-1} + \gamma\tilde{\eta}_{2t}$; (v) $[\tilde{f}_t,\ \tilde{u}_t,\ \tilde{\eta}_{1t},\ \tilde{\eta}_{2t}]' \sim i.i.d.\ N(\mathbf{0}, \mathbf{I})$. The results in the table are based on the following parameter settings: $\psi = 10$, $\lambda = 0.5$, $\omega = 10$, $\delta_1 = 0$, $\delta_2 = 0$, and $\gamma = \sqrt{1-\phi^2}$. The table is constructed by estimating the regression

$$\sum_{i=1}^{k}\tilde{r}_{t+i} = \alpha_k + \beta_{1k}\tilde{z}_{1t} + \beta_{2k}\tilde{z}_{2t} + \tilde{\varepsilon}_{t+k},$$

for $k = 1$, $k = 2$, $k = 3$, and $k = 4$. The one-period returns are nonoverlapping, and the two-, three-, and four-period returns are overlapping one-period observations. There are 61 one-period returns, 60 two-period returns, 59 three-period returns, and 58 four-period returns. The t-ratios are corrected for heteroscedasticity and (for multiperiod returns) the autocorrelation of overlapping residuals using the methods of Hansen (1982) and White (1980). Each line of the table corresponds to a simulation with 10,000 trials.

iting distributions. For the sample R^2 and the t-ratio that imposes the null hypothesis, this is in fact the case.[11] The conventional t-ratio, however, tends to reject the null too often in these regressions.

In light of the fact that the t-ratio that imposes the null performs well in the simulations, it is not at all surprising to find that the conventional t-ratio is biased toward rejecting the null. The sample variance of the regression residuals is always less than or equal to the sample variance of the returns. It follows, therefore, that the standard errors that are obtained by imposing the null will, on average, be larger than the standard errors that are computed in the conventional manner. This in turn implies that the conventional t-ratio will generally be larger than the t-ratio that is computed by imposing the null.

The bias in the case of one-period returns does not appear to be all that severe. It gets progressively worse, though, as we move to returns measured over longer horizons. In simulations where the autocorrelation coefficient for the instruments is set equal to 0.9, the regressions for four-period returns yield a 97.5% cutoff for the conventional t-ratio of 3.14. This is much larger that the nominal cutoff of 1.96. It seems that in regressions where the instruments are highly persistent, the standard errors used to construct the conventional t-ratio tend to underestimate the actual variation in the OLS estimator of the slope coefficients. Consequently the conventional t-ratio may be large even if returns are unpredictable.

The t-ratio that imposes the null, on the other hand, does not suffer from this problem. It performs reasonably well regardless of the return horizon. Thus, imposing the null does seem to improve the size of the tests in the long-horizon setting. The most interesting feature of Table 7, however, is the shift in the distribution of the sample R^2 as we move to longer return horizons. Notice how the mean of the sample R^2 behaves as the return horizon gets longer. There is little change if the instruments are serially uncorrelated, but the mean increases in a monotonic fashion if the instruments are autocorrelated. With $\phi = 0.9$, the mean of the sample R^2 for the four-period regressions is 10.9%. This is over three times as large as the corresponding value for the one-period regressions.

There is also a substantial increase in the standard deviation of the sample R^2 as the return horizon gets longer. As a result, the probability of observing large values of the sample R^2 is much greater in the multiperiod regressions than in the one-period regressions. With ϕ set equal to 0.9, the 95% cutoff for the sample R^2 is 9.9% for the

[11] According to Theorem 1.2, the sample R^2 should have a mean of $2/61 = 3.3\%$, a standard deviation of $2/61 = 3.3\%$, and 95% of the values should fall below $5.99/61 = 9.8\%$.

one-period regressions. The corresponding value for the four-period regressions is 29.6%. This shows that it is quite possible for a long-horizon regression to yield large values of the sample R^2 even though the returns are in reality unpredictable.

4.4 The evidence on power

Although the simulations of Section 4.3 confirm that large values of the conventional t-ratio and sample R^2 are not necessarily an indication that long-horizon returns are highly predictable, it does not appear that they can explain all of the results in Tables 2 and 3. The simulations indicate that the conventional t-ratio has reasonable size in long-horizon regressions that use nonoverlapping data. This means that poor size is not a likely explanation for the large t-ratios reported for the annual regressions in Tables 2 and 3. Nevertheless, large t-ratios for the one-year returns do not necessarily indicate that they are highly predictable. If the conventional t-ratio has high power, then it is conceivable that a small amount of predictability might generate the observed results.

Evidence on the power of the t-ratios is presented in Tables 8 and 9. Table 8 examines the distributional properties of both the conventional t-ratio and the t-ratio that imposes the null for the situation where returns are slightly predictable. More specifically, this table reports the mean, standard deviation, and percentage points for the t-ratios in long-horizon regressions that have a population R^2 of 5%. The results in panel A are for one-period returns. Those in panel B are for four-period returns. Each panel presents results for two different settings of δ_1. This is done because different settings for this parameter generate different degrees of conditional heteroscedasticity.

The results in Table 8 reveal that the conventional t-ratio does have a power advantage in long-horizon regressions. With $\delta_1 = 5$, for instance, panel A indicates that the mean of the conventional t-ratio ranges from a low of 1.45 to a high of 1.76. The corresponding values for the t-ratio that imposes the null are 1.29 and 1.53. Similar results are obtained when $\delta_1 = 10$. The mean of the conventional t-ratio varies from 1.39 to 1.58. For the t-ratio that imposes the null it varies from 1.23 to 1.36. These results suggest that the distributional properties of the t-ratios are not overly sensitive to the degree-conditional heteroscedasticity displayed by returns.

The power advantage of the conventional t-ratio is more pronounced in the four-period regressions. With $\delta_1 = 5$, the mean ranges from 1.02 to 1.76 for the conventional t-ratio, and from 0.63 to 1.20 for the t-ratio that imposes the null. Because the conventional t-ratio demonstrates good power in the long-horizon setting, it is possible for a regression with a small population R^2 to produce what might be

Table 8
Percentage points for the slope t-ratios in simulations where the population R^2 is 5%

Horizon	δ_1	λ	ϕ	Mean	SD	0.50	0.90	0.95	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
Panel A				Conventional t-ratio (H_0: $\beta_{1k} = 0$)							t-ratio that imposes H_0: $\beta_{1k} = 0$						
1-period	5	0.688	0.3	1.76	1.18	1.72	3.27	3.77	4.23	4.79	1.53	0.92	1.59	2.66	2.94	3.18	3.41
1-period	5	0.688	0.6	1.72	1.18	1.69	3.22	3.74	4.22	4.73	1.50	0.92	1.56	2.64	2.92	3.16	3.45
1-period	5	0.688	0.9	1.45	1.18	1.43	2.97	3.46	3.90	4.48	1.29	0.98	1.36	2.51	2.82	3.06	3.33
1-period	10	0.397	0.3	1.58	1.20	1.54	3.14	3.64	4.07	4.60	1.36	0.94	1.44	2.54	2.80	3.04	3.28
1-period	10	0.397	0.6	1.56	1.19	1.52	3.09	3.61	4.07	4.61	1.35	0.94	1.43	2.50	2.79	3.04	3.29
1-period	10	0.397	0.9	1.39	1.19	1.36	2.90	3.41	3.85	4.43	1.23	0.98	1.29	2.43	2.75	3.02	3.27
Panel B				Conventional t-ratio (H_0: $\beta_{1k} = 0$)							t-ratio that imposes H_0: $\beta_{1k} = 0$						
4-periods	5	1.030	0.3	1.76	1.56	1.64	3.49	4.24	4.95	6.04	1.20	0.80	1.34	2.06	2.23	2.36	2.50
4-periods	5	0.650	0.6	1.43	1.59	1.34	3.33	4.08	4.80	5.68	0.94	0.88	1.11	1.92	2.09	2.22	2.36
4-periods	5	0.402	0.9	1.02	1.66	0.93	2.26	3.80	4.60	5.51	0.63	0.96	0.81	1.75	1.92	2.05	2.22

Table 8
(continued)

Horizon	δ_1	λ	ϕ	Mean	SD	0.50	0.90	0.95	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
4-periods	10	0.595	0.3	1.73	1.57	1.63	3.51	4.20	4.98	6.13	1.17	0.82	1.34	2.05	2.22	2.36	2.51
4-periods	10	0.375	0.6	1.41	1.62	1.32	3.34	4.15	4.81	5.92	0.91	0.90	1.11	1.91	2.07	2.20	2.35
4-periods	10	0.232	0.9	1.03	1.70	0.92	3.07	3.89	4.68	5.80	0.63	0.96	0.80	1.74	1.91	2.05	2.21
Normal (0,1)				0.00	1.00	0.00	1.28	1.65	1.96	2.33	0.00	1.00	0.00	1.28	1.65	1.96	2.33

The data generating process for excess returns is a conditional linear factor model of the form

$$\tilde{r}_t = m_t + b_t \tilde{f}_t + \psi \tilde{u}_t,$$

where (i) $m_t = \lambda b_t$; (ii) $b_t = \omega + \delta_1 \tilde{z}_{1,t-1} + \delta_2 \tilde{z}_{2,t-1}$; (iii) $\tilde{z}_{1t} = \phi \tilde{z}_{1,t-1} + \gamma \tilde{\eta}_{1t}$; (iv) $\tilde{z}_{2t} = \phi \tilde{z}_{2,t-1} + \gamma \tilde{\eta}_{2t}$; (v) $[\tilde{f}_t,\ \tilde{u}_t,\ \tilde{\eta}_{1t},\ \tilde{\eta}_{2t}]' \sim i.i.d.\ N(\mathbf{0}, \mathbf{I})$. The results in the table are based on the following parameter settings: $\psi = 10$, $\omega = 10$, $\delta_1 = 0$, and $\gamma = \sqrt{1-\phi^2}$. The table is constructed by estimating the regression

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} \tilde{z}_{1t} + \beta_{2k} \tilde{z}_{2t} + \tilde{\varepsilon}_{t+k},$$

for $k = 1$ and $k = 4$. The one-period returns are nonoverlapping, and the four-period returns are overlapping one-period observations. There are 61 one-period returns and 58 four-period returns. In each case, λ is set so that the population value of the R^2 for the regression is 5%. The t-ratios are corrected for heteroscedasticity and (for four-period returns) the autocorrelation of overlapping residuals using the methods of Hansen (1982) and White (1980). Each line of the table corresponds to a simulation with 10,000 trials.

Table 9
Percentage points for the sample R^2 and associated t-ratios in simulations with a population R^2 of 5%

Horizon	δ_1	λ	ϕ	Mean	SD	0.01	0.025	0.50	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
Panel A				Variance-stabilized t-ratio (H_0: $\rho^2 = 5\%$)							Sample R^2 for the regression						
1-period	5	0.688	0.3	0.36	1.01	−1.62	−1.39	0.31	2.50	3.00	8.2	6.4	6.8	17.0	20.8	24.3	28.4
1-period	5	0.688	0.6	0.34	1.00	−1.61	−1.39	0.29	2.47	3.01	8.1	6.4	6.7	16.8	20.6	24.2	28.5
1-period	5	0.688	0.9	0.16	1.00	−1.78	−1.55	0.09	2.32	2.79	7.0	6.0	5.5	15.2	19.0	22.3	26.5
1-period	10	0.397	0.3	0.38	1.02	−1.56	−1.35	0.30	2.61	3.11	8.8	7.3	7.0	19.0	23.2	27.0	31.9
1-period	10	0.397	0.6	0.36	1.02	−1.60	−1.38	0.30	2.62	3.11	8.6	7.2	6.9	18.4	22.9	27.0	31.7
1-period	10	0.397	0.9	0.21	1.02	−1.76	−1.52	0.14	2.41	2.99	7.6	6.6	5.8	16.7	20.8	25.0	30.0
Panel B				Variance-stabilized t-ratio (H_0: $\rho^2 = 5\%$)							Sample R^2 for the regression						
4-periods	5	1.030	0.3	0.47	1.01	−1.48	−1.26	0.39	2.64	3.11	9.3	7.4	7.6	19.7	23.9	27.5	31.8
4-periods	5	0.650	0.6	0.63	1.07	−1.34	−1.10	0.51	2.99	3.59	11.4	9.4	9.1	24.8	30.3	34.9	40.5
4-periods	5	0.402	0.9	0.78	1.12	−1.19	−0.97	0.64	3.33	4.03	13.7	11.4	10.8	30.3	36.4	42.3	48.9

Table 9
(continued)

Horizon	δ_1	λ	ϕ	Mean	SD	0.01	0.025	0.50	0.975	0.99	Mean	SD	0.50	0.90	0.95	0.975	0.99
4-periods	10	0.595	0.3	0.48	1.02	−1.47	−1.24	0.40	2.69	3.19	9.5	7.6	7.7	20.2	24.2	28.2	33.2
4-periods	10	0.375	0.6	0.66	1.09	−1.32	−1.09	0.55	3.07	3.73	11.8	9.8	9.4	25.9	31.4	36.0	41.3
4-periods	10	0.232	0.9	0.83	1.16	−1.17	−0.95	0.67	3.51	4.18	14.4	12.0	11.2	31.7	38.8	44.5	50.6
Normal (0,1)				0.00	1.00	−2.33	−1.96	0.00	1.96	2.33							

The data generating process for excess returns is a conditional linear factor model of the form

$$\tilde{r}_t = m_t + b_t \tilde{f}_t + \psi \tilde{u}_t,$$

where (i) $m_t = \lambda b_t$; (ii) $b_t = \omega + \delta_1 \tilde{z}_{1,t-1} + \delta_2 \tilde{z}_{2,t-1}$; (iii) $\tilde{z}_{1t} = \phi \tilde{z}_{1,t-1} + \gamma \tilde{\eta}_{1t}$; (iv) $\tilde{z}_{2t} = \phi \tilde{z}_{2,t-1} + \gamma \tilde{\eta}_{2t}$; (v) $[\tilde{f}_t,\ \tilde{u}_t,\ \tilde{\eta}_{1t},\ \tilde{\eta}_{2t}]' \sim i.i.d.\ N(\mathbf{0}, \mathbf{I})$. The results in the table are based on the following parameter settings: $\psi = 10$, $\omega = 10$, $\delta_2 = 0$, and $\gamma = \sqrt{1 - \phi^2}$. The table is constructed by estimating the regression

$$\sum_{i=1}^{k} \tilde{r}_{t+i} = \alpha_k + \beta_{1k} \tilde{z}_{1t} + \beta_{2k} \tilde{z}_{2t} + \tilde{\varepsilon}_{t+k},$$

for $k = 1$ and $k = 4$. The one-period returns are nonoverlapping, and the four-period returns are overlapping one-period observations. There are 61 one-period returns and 58 four-period returns. In each case, λ is set so that the population value of the R^2 for the regression is 5%. The t-ratios, which are constructed using a variance-stabilizing transformation of the limiting distribution of the sample multiple correlation coefficient, are corrected for heteroscedasticity and autocorrelation using the procedures outlined in Andrews (1991). Each line in the table corresponds to a simulation with 10,000 trials.

interpreted as evidence that returns are highly predictable. The 97.5% cutoff for the conventional t-ratio is typically above 4.00, and the 99% cutoff can exceed 6.00. These values are in the same range as those reported in Tables 2 and 3 for the annual regressions.

Now consider the distributional properties of the sample R^2. Table 9 reports the mean, standard deviation, and percentage points for both the sample R^2 and the t-ratio associated with the sample multiple correlation coefficient. This t-ratio is computed using the variance-stabilizing transformation discussed in Section 1.4. The results in panel A are for one-period returns; those in panel B are for four-period returns. The first aspect to note about these results is that the sample R^2 is biased upward. In panel A the mean of the sample R^2 ranges from 7.0% to 8.8%, and in panel B it ranges from 9.3% to 14.4%. The increase in the bias for the regressions that use overlapping returns is consistent with the implications of the distributional theory for the sample R^2.

Not surprisingly, the upward bias in the sample R^2 is reflected in the distribution of the t-ratio for the multiple correlation coefficient. The mean of this distribution is always greater than zero, and its variance is always greater than one. It appears from these results that the t-ratio for the multiple correlation coefficient does not possess good size in situations where the population R^2 is close to zero. This finding is reasonably intuitive. The sample R^2 can never be less than zero. It follows, therefore, that when the population R^2 is close to zero, the asymptotic normal approximation is not likely to perform well in small samples.

There is also a substantial degree of variation in the value of the sample R^2 for the regressions. It is not uncommon to see a large sample R^2 even though the population R^2 is only 5%. In the one-period regressions, the 95% cutoff for the sample R^2 hovers around 20%. For the four-period regressions, however, it goes as high as 38.8%. This means that under certain circumstances we would expect to observe a sample R^2 that is greater than 38.8% in 5 out of 100 regressions, despite the fact that the population R^2 is only 5%. Thus, the simulations indicate that in many cases it not possible to conclude from large values of the sample R^2 that long-horizon returns are highly predictable. The long-horizon estimates are simply too imprecise to be able to draw firm inferences.

4.5 Implications

The Monte Carlo experiments clearly illustrate the effects that small sample sizes and autocorrelated errors can have on the distributional properties of the t-ratios and sample R^2. The conventional t-ratio displays poor size under the null, but good power under the alternative.

It is not uncommon, therefore, to observe large realizations of this t-ratio in regressions where the population R^2 is either small or zero. The t-ratio that imposes the null, on the other hand, displays good size under the null, but is not as powerful as the conventional t-ratio under the alternative. This disparity in power may explain in large part why the conventional t-ratio rejects the null so often when the t-ratio that imposes the null does not. Evidence against the null, however, is not synonymous with strong predictability at long-horizon horizons.

The most telling evidence from the Monte Carlo experiments concern the distributional properties of the sample R^2. The simulations reveal that long-horizon regressions can easily yield large values of the sample R^2 under circumstances where returns do not display a high degree of predictability. This is consistent with the theoretical analysis, which implies that we would expect the sample R^2 to increase as the horizon gets longer, and the empirical results, which indicate that the standard error of the sample R^2 increases dramatically at long horizons. On balance, the results are consistent with the view that the increase in the sample R^2 at long horizons is driven more by statistical considerations than by economic forces.

These results should carry over to studies that rely on other methodologies as well. Campbell and Shiller (1988), for example, use a vector autoregressive (VAR) approach to examine the relation between dividend growth rates, dividend yields, and earnings:price ratios. First they apply a log transformation to annual data on the dividend yield and earnings:price ratio. Then they estimate VAR specifications using observations for the 1901–1987 time period. The dividend yields and dividend growth rates used to estimate the models are nonoverlapping, but the earnings:price ratios are computed based on a 30-year moving average of earnings.

The VAR results of Campbell and Shiller (1988) seem to indicate that the log dividend yield and log earnings:price ratio have strong forecasting power for dividend growth. The t-ratios for these two variables are −6.23 and 4.54, respectively, and the sample R^2 for the model is 36.1%. But there may be good reason to regard these results as somewhat suspect. The VAR estimates of the first-order autoregressive coefficients for the dividend growth rate, log dividend yield, and log earnings:price ratio are 0.33, 0.61, and 0.87, respectively. If these estimates accurately reflect the degree of autocorrelation in the data, then the mechanism documented in the simulations is most likely at work in these VAR models.

The analysis also has implications for studies besides those that focus strictly on the ability to predict returns. Fama (1990), for example, uses a regression approach to examine the relation between stock returns and future production growth rates. He finds that future production growth rates explain 6% of the variation in monthly returns on the NYSE value-weighted index. But the proportion rises to 43% for annual returns. Fama (1990) argues, therefore, that real activity explains a larger fraction of return variation at long horizons. Again, however, this inference may not be warranted. His regression model has 137 annual observations, uses overlapping returns observed at quarterly intervals, and employs eight explanatory variables, each of which has an estimated first-order autocorrelation coefficient of 0.33. Thus, the sample R^2 for the regression almost certainly has a large standard error.

5. Conclusions

Recent studies by Campbell and Shiller (1988) and Fama and French (1989) contend that long-horizon stock and bond returns are highly predictable. The primary basis for this claim seems to be that the sample R^2 from predictive regressions increases with the length of the return horizon. Fama and French (1989), for instance, find that the sample R^2 increases from around 3% for monthly returns to well over 25% for four-year returns. They attribute this increase in the sample R^2 to strong predictability at long horizons that can be traced to persistence is short-horizon expected returns. As a result, they argue that the predictability of long-horizon returns arises as a natural consequence of changing business conditions.

This article offers a different view of the long-horizon evidence. Long-horizon studies typically measure predictability using regression models that, more often than not, employ overlapping returns. The overlapping returns produce serially correlated errors that, along with small sample sizes, make it difficult to draw precise inferences. Long-horizon regressions can yield large values of the sample R^2 in situations where the population R^2 is small or zero. Moreover, long-horizon regressions with a small or zero population R^2 can produce t-ratios that might be interpreted as evidence of strong predictability. Consequently the conclusion that long-horizon returns are highly predictable does not appear to be justified. The results suggest that the increase in the sample R^2 at long horizons has more to do with statistical properties of this estimator than with changing business conditions.

Appendix

The proofs of Theorems 1.1–1.4 assume a working knowledge of the asymptotic theory associated with the generalized method of moments. See Hansen (1982) for an detailed discussion of the GMM procedure.

Proof of Theorem 1.1. Start with the following disturbance vector:

$$h(\tilde{x}_t, \theta) = \begin{pmatrix} \sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k \\ (\sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k)\tilde{z}_t \end{pmatrix} \tag{1A}$$

where $\tilde{x}_t \equiv [\sum_{i=1}^{k} \tilde{r}_{t+i},\ \tilde{z}_t']'$ and $\theta \equiv [\alpha_k, \beta_k']'$. Because the system in Equation (1A) is exactly identified, it is readily apparent that the GMM estimator is obtained by replacing each element of the vector θ with its sample analog. The sample analog of the vector β_k is the least-squares estimator $\hat{\beta}_k$. Thus, the asymptotic results given by Hansen (1982),

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, (D'S^{-1}D)^{-1}), \tag{2A}$$

where

$$D = E\left[\frac{\partial h(\tilde{x}_t, \theta)}{\partial \theta'}\right] \quad \text{and} \quad S = \sum_{j=-\infty}^{\infty} E[h(\tilde{x}_t, \theta)h(\tilde{x}_{t-j}, \theta)'] \tag{3A}$$

can be used to derive the limiting distribution of $\hat{\beta}_k$.

An analytic version of D can easily be obtained by taking the expected value of the Jacobian of $h(\tilde{x}_t, \theta)$ with respect to θ. The resulting expression is

$$D = \begin{bmatrix} -1 & -\mu_z' \\ -\mu_z & -E[\tilde{z}_t\tilde{z}_t'] \end{bmatrix}. \tag{4A}$$

Applying standard results on the inverse of a partitioned matrix yields

$$D^{-1} = \begin{bmatrix} -(1 + \mu_z'\Sigma_{zz}^{-1}\mu_z) & \mu_z'\Sigma_{zz}^{-1} \\ \Sigma_{zz}^{-1}\mu_z & -\Sigma_{zz}^{-1} \end{bmatrix}, \tag{5A}$$

where $\Sigma_{zz} \equiv E[(\tilde{z}_t - \mu_z)(\tilde{z}_t - \mu_z)']$ denotes the variance-covariance matrix of the instrumental variables. Using this result, the lower-right $m \times m$ submatrix of $D^{-1}SD^{-1\prime}$ can be written as

$$V \equiv \sum_{j=-\infty}^{\infty} E\left(\begin{bmatrix} \mu_z'\Sigma_{zz}^{-1} \\ -\Sigma_{zz}^{-1} \end{bmatrix}' \begin{bmatrix} \tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j} & \tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j}\tilde{z}_{t-j}' \\ \tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j}\tilde{z}_t & \tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j}\tilde{z}_t\tilde{z}_{t-j}' \end{bmatrix} \begin{bmatrix} \mu_z'\Sigma_{zz}^{-1} \\ -\Sigma_{zz}^{-1} \end{bmatrix}\right), \tag{6A}$$

where $\tilde{\varepsilon}_{t+k} \equiv \sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k$. Therefore, the limiting distribution of the least-squares estimator is

$$\sqrt{T}(\hat{\beta}_k - \beta_k) \xrightarrow{d} N(\mathbf{0}, V), \tag{7A}$$

with V given by

$$V = \Sigma_{zz}^{-1}\left[\sum_{j=-\infty}^{\infty} E\left[\tilde{\varepsilon}_{t+k}\tilde{\varepsilon}_{t+k-j}(\tilde{z}_t - \mu_z)(\tilde{z}_{t-j} - \mu_z)'\right]\right]\Sigma_{zz}^{-1}. \quad (8A)$$

Proof of Theorem 1.2. Let $\tilde{\eta}_{t+k}$ denote the disturbance vector $\tilde{\varepsilon}_{t+k}(\tilde{z}_t - \mu_z)$. Now suppose that $\tilde{\eta}_{t+k}$ is serially uncorrelated and β_k is equal to zero. The matrix V can be written as

$$V=\Sigma_{zz}^{-1}\left[\sigma_k^2\Sigma_{zz}+\text{cov}\left(\left(\sum_{i=1}^{k}\tilde{r}_{t+i}-\mu_k\right)^2, (\tilde{z}_t - \mu_z)(\tilde{z}_t - \mu_z)'\right)\right]\Sigma_{zz}^{-1}. \quad (9A)$$

If returns are conditionally homoscedastic, then the covariance term in Equation (9A) disappears and the limiting distribution of the least-squares estimator becomes

$$\sqrt{T}\hat{\beta}_k \xrightarrow{d} N(\mathbf{0}, \sigma_k^2\Sigma_{zz}^{-1}). \quad (10A)$$

Therefore, the Wald statistic,

$$\hat{W}_D \equiv T\left(\frac{\hat{\beta}_k'\hat{\Sigma}_{zz}\hat{\beta}_k}{\hat{\sigma}_k^2}\right), \quad (11A)$$

converges to a chi-square random variable with m degrees of freedom. The right-hand side of Equation (11A) is equal to TR_k^2.

Proof of Theorem 1.3. Suppose the vector of slope coefficients β_k is equal to zero. Under these circumstances, the limiting distribution of the least-squares estimator is

$$\sqrt{T}\hat{\beta}_k \xrightarrow{d} N(\mathbf{0}, V), \quad (12A)$$

with V given by Equation (8A). Premultiply the distribution shown in (12A) by the $2m \times m$ vector $[I,\ \hat{\Sigma}_{zz}\hat{\sigma}_k^{-2}]'$ to obtain

$$\sqrt{T}\begin{pmatrix} \hat{\beta}_k \\ \hat{\sigma}_{zr}\hat{\sigma}_k^{-2} \end{pmatrix} \xrightarrow{d} SN\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} V & V\Sigma_{zz}\hat{\sigma}_k^{-2} \\ \Sigma_{zz}V\hat{\sigma}_k^{-2} & \Sigma_{zz}V\Sigma_{zz}\hat{\sigma}_k^{-4} \end{bmatrix}\right), \quad (13A)$$

where $SN(\cdot)$ denotes a singular normal distribution. It is apparent that TR_k^2 can be written as:

$$TR_k^2 = \frac{T}{2}\begin{pmatrix} \hat{\beta}_k \\ \hat{\sigma}_{zr}\hat{\sigma}_k^{-2} \end{pmatrix}'\begin{pmatrix} \mathbf{0} & I \\ I & \mathbf{0} \end{pmatrix}\begin{pmatrix} \hat{\beta}_k \\ \hat{\sigma}_{zr}\hat{\sigma}_k^{-2} \end{pmatrix}. \quad (14A)$$

Thus, by Slutsky's theorem, the limiting distribution of TR_k^2 is that of a quadratic form in a singular normal random vector. Searle (1971) gives formulas for the first two central moments of such quadratic forms. In particular, he shows that if

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim SN\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\right), \tag{15A}$$

the expected value and variance of the quadratic form

$$q \equiv \frac{1}{2}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}'\begin{pmatrix} \mathbf{0} & A \\ A & \mathbf{0} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{16A}$$

are given by

$$\mu_q = tr(AC_{21}) \quad \text{and} \quad \sigma_q^2 = tr(AC_{21})^2 + tr(AC_{22}AC_{11}). \tag{17A}$$

Applying these formulas to the distribution shown in Equation (13A) yields the desired results.

Proof of Theorem 1.4. Under the alternative hypothesis the population R^2 lies between zero and one. It can therefore be estimated in standard fashion using the econometric specification

$$h(\tilde{x}_t, \theta) = \begin{pmatrix} \sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k \\ (\sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k)\tilde{z}_t \\ \sum_{i=1}^{k} \tilde{r}_{t+i} - \mu_k \\ (1-\rho_k^2)(\sum_{i=1}^{k} \tilde{r}_{t+i} - \mu_k)^2 - (\sum_{i=1}^{k} \tilde{r}_{t+i} - \alpha_k - \tilde{z}_t'\beta_k)^2 \end{pmatrix}, \tag{18A}$$

where $\tilde{x}_t \equiv [\sum_{i=1}^{k} \tilde{r}_{t+i},\ \tilde{z}_t']'$ and $\theta \equiv [\mu_k,\ \alpha_k,\ \beta_k',\ \rho_k^2]'$. The system in Equation (18A) is exactly identified, so the GMM estimator is again obtained by replacing each element of the vector θ with its sample analog. Taking the expected value of the Jacobian of $h(\tilde{x}_t, \theta)$ with respect to θ yields a block diagonal matrix. The first block is given by Equation (4A), and the second is simply diag$\{-1,\ -\sigma_k^2\}$. Because the D matrix takes this form, only the lower right-most element of S will have an impact on the distribution of R_k^2. This element of S is given by

$$S_{\rho^2} \equiv \sum_{j=-\infty}^{\infty} E[\tilde{\xi}_{t+k}\tilde{\xi}_{t+k-j}], \tag{19A}$$

where $\tilde{\xi}_{t+k} = ((1-\rho_k^2)(\sum_{i=1}^{k} \tilde{r}_{t+i} - \mu_k)^2 - \tilde{\varepsilon}_{t+k}^2)$ and $\tilde{\varepsilon}_{t+k} \equiv (\sum_{i=1}^{k} \tilde{r}_{t+i} -$

$\alpha_k - \tilde{z}'_t \beta_k)$. Thus the limiting distribution of R_k^2 is

$$\sqrt{T}(R_k^2 - \rho_k^2) \xrightarrow{d} N(0, \sigma_R^2), \tag{20A}$$

with σ_R^2 given by

$$\sigma_R^2 \equiv \sum_{j=-\infty}^{\infty} E\left[\frac{\tilde{\xi}_{t+k}\tilde{\xi}_{t+k-j}}{\sigma_k^4}\right]. \tag{21A}$$

The Monte Carlo experiment

This section outlines the derivation of the population value of R^2 for a k-period return horizon. The data generating process for excess returns takes the form of a conditional linear factor model:

$$\begin{aligned}
\tilde{r}_t &= m_t + b_t \tilde{f}_t + \psi \tilde{u}_t & \psi > 0 \\
\tilde{z}_{it} &= \phi \tilde{z}_{i,t-1} + \gamma \tilde{\eta}_{it} & \gamma > 0,\ -1 < \phi < 1,\ i = 1, 2 \\
m_t &= \lambda b_t & \lambda > 0 \\
b_t &= \omega + \delta_1 \tilde{z}_{1,t-1} + \delta_2 \tilde{z}_{2,t-1} & \omega,\ \delta_1,\ \delta_2 \geq 0
\end{aligned} \tag{22A}$$

where

$$[\tilde{f}_t,\ \tilde{u}_t,\ \tilde{\eta}_{1t},\ \tilde{\eta}_{2t}]' \sim i.i.d.\ N(\mathbf{0},\ \mathbf{I}).$$

First, set $\delta_2 = 0$ and note the following results:

$$\begin{aligned}
\tilde{z}_{1t} &= \gamma \textstyle\sum_{i=0}^{\infty} \phi^i \tilde{\eta}_{1,t-i} \\
\tilde{z}_{1t} &\sim N\left(0, \gamma^2/(1-\phi^2)\right) \\
\operatorname{cov}(\textstyle\sum_{i=1}^{k} \tilde{r}_{t+i}, \tilde{z}_{1t}) &= \lambda \delta_1 \operatorname{var}(\tilde{z}_{1t}) \textstyle\sum_{i=1}^{k} \phi^{i-1}.
\end{aligned} \tag{23A}$$

Next, consider the general expression for the population value of R^2 for a k-period return horizon:

$$\rho_k^2 \equiv \left(\frac{\operatorname{cov}(\sum_{i=1}^{k} \tilde{r}_{t+i}, \tilde{z}_{1t})^2}{\operatorname{var}(\sum_{i=1}^{k} \tilde{r}_{t+i}) \operatorname{var}(\tilde{z}_{1t})}\right). \tag{24A}$$

The quantity $\sum_{i=1}^{k} \phi^{i-1}$ is equal to $(1-\phi^k)/(1-\phi)$, so Equation (24A) can be written as

$$\rho_k^2 = \rho_1^2 \left(\frac{1-\phi^k}{1-\phi}\right)^2 \left(\frac{\operatorname{var}(\sum_{i=1}^{k} \tilde{r}_{t+i})}{\operatorname{var}(\tilde{r}_{t+1})}\right)^{-1}, \tag{25A}$$

where

$$\rho_1^2 \equiv \left(\frac{\lambda^2 \delta_1^2 \operatorname{var}(\tilde{z}_{1t})}{\operatorname{var}(\tilde{r}_{t+1})} \right) \tag{26A}$$

denotes the population value of R^2 for the one-period return horizon. But $\operatorname{var}(\tilde{r}_{t+1})$ is given by

$$\begin{aligned} \operatorname{var}(\tilde{r}_{t+1}) &= E[(\lambda\delta_1 \tilde{z}_{1t} + (\omega + \delta_1 \tilde{z}_{1t})\tilde{f}_{t+1} + \psi \tilde{u}_{t+1})^2] \\ &= \psi^2 + \omega^2 + \delta_1^2(1+\lambda^2)\operatorname{var}(\tilde{z}_{1t}) \\ &= \left(\frac{(\psi^2+\omega^2)(1-\phi^2) + \gamma^2\delta_1^2(1+\lambda^2)}{1-\phi^2} \right), \end{aligned} \tag{27A}$$

so Equation (26A) becomes

$$\rho_1^2 = \left(\frac{\lambda^2 \delta_1^2 \gamma^2}{(\psi^2+\omega^2)(1-\phi^2) + \gamma^2\delta_1^2(1+\lambda^2)} \right). \tag{28A}$$

Now consider the k-period variance ratio. The general formula given by Lo and MacKinlay (1988) and Richardson and Smith (1991) is

$$\frac{\operatorname{var}(\sum_{i=1}^{k} \tilde{r}_{t+i})}{\operatorname{var}(\tilde{r}_{t+1})} = k + 2\sum_{i=1}^{k-1}(k-i)\operatorname{corr}(\tilde{r}_{t+k}, \tilde{r}_{t+k-i}), \tag{29A}$$

where $\operatorname{corr}(\cdot)$ denotes the correlation operator. To find the autocorrelation coefficient at lag j note that $\tilde{r}_{t+j}$ can be written as

$$\begin{aligned} \tilde{r}_{t+j} &= \lambda\omega + \delta_1(\lambda + \tilde{f}_{t+j})\left(\phi^j \tilde{z}_{1,t-1} + \gamma \sum_{i=1}^{j} \phi^{j-i} \tilde{\eta}_{1,t+i-1} \right) \\ &\quad + \omega \tilde{f}_{t+j} + \psi \tilde{u}_{t+j}. \end{aligned} \tag{30A}$$

Thus the covariance between $\tilde{r}_t$ and $\tilde{r}_{t+j}$ is given by

$$\operatorname{cov}(\tilde{r}_t, \tilde{r}_{t+j}) = \lambda^2 \delta_1^2 \phi^j \operatorname{var}(\tilde{z}_{1,t-1}). \tag{31A}$$

Equations (27A) and (31A) can be combined to yield the formula for the autocorrelation coefficient at lag j:

$$\operatorname{corr}(\tilde{r}_t, \tilde{r}_{t+j}) = \left(\frac{\lambda^2 \delta_1^2 \gamma^2 \phi^j}{(\psi^2+\omega^2)(1-\phi^2) + \gamma^2\delta_1^2(1+\lambda^2)} \right). \tag{32A}$$

Therefore, Equation (24A) becomes:

$$\rho_k^2 = \left(\frac{\left(\frac{1-\phi^k}{1-\phi}\right)^2 \left(\frac{\lambda^2\delta_1^2\gamma^2}{(\psi^2+\omega^2)(1-\phi^2)+\gamma^2\delta_1^2(1+\lambda^2)}\right)}{k+2\left(\frac{\lambda^2\delta_1^2\gamma^2}{(\psi^2+\omega^2)(1-\phi^2)+\gamma^2\delta_1^2(1+\lambda^2)}\right)\sum_{i=1}^{k-1}(k-i)\phi^i} \right). \tag{33A}$$

References

Anderson, T., 1984, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York.

Andrews, D. W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Bekaert, G., and R. Hodrick, 1992, "Characterizing Predictable Components in Equity and Foreign Exchange Rates of Return," *Journal of Finance*, 47, 467–509.

Campbell, J., 1987, "Stock Returns and the Term Structure," *Journal of Financial Economics*, 18, 373–399.

Campbell, J., and R. Shiller, 1988, "Stock Prices, Earnings and Expected Dividends," *Journal of Finance*, 46, 661–676.

Fama, E., 1990, "Stock Returns, Expected Returns, and Real Activity," *Journal of Finance*, 45, 1089–1108.

Fama, E., and R. Bliss, 1987, "The Information in Long-Maturity Forward Rates," *American Economic Review*, 77, 680–692.

Fama, E., and K. French, 1988a, "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 19, 3–29.

Fama, E., and K. French, 1988b, "Permanent and Temporary Components of Stock Prices," *Journal of Political Economy*, 96, 246–273.

Fama, E., and K. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics*, 25, 23–49.

Fama, E., and W. Schwert, 1977, "Asset Returns and Inflation," *Journal of Financial Economics*, 5, 115–146.

Ferson, W., and C. Harvey, 1991, "The Variation in Economic Risk Premiums," *Journal of Political Economy* 99, 385–415.

Foster, D., T. Smith, and R. Whaley, 1996, "Assessing the Goodness-of-Fit of Asset Pricing Models: The Distribution of the Maximal R^2," working paper, Fuqua School of Business, Duke University; forthcoming in *Journal of Finance*.

Goetzmann, W., and P. Jorion, 1993, "Testing the Predictive Power of Dividend Yields," *Journal of Finance*, 48, 663–679.

Goetzmann, W., and P. Jorion, 1995, "A Longer Look at Dividend Yields," *Journal of Business*, 68, 483–508.

Granger, C., and P. Newbold, 1974, "Spurious Regressions in Econometrics," *Journal of Econometrics*, 2, 111–120.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hodrick, R., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

Jegadeesh, N., 1991, "Seasonality in Stock Price Mean Reversion: Evidence from the U.S. and U.K.," *Journal of Finance*, 46, 1427–1444.

Keim, D., and R. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics*, 17, 357–390.

Kim, M., C. Nelson, and R. Startz, 1991, "Mean Reversion in Stock Prices: A Reappraisal of the Empirical Evidence," *Review of Economic Studies*, 58, 515–528.

Lo, A., and C. MacKinlay, 1988, "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test," *Review of Financial Studies*, 1, 41–66.

Lo, A., and C. MacKinlay, 1995, "Maximizing Predictability in the Stock and Bond Markets," working paper, Sloan School of Management, MIT.

Nelson, C., and M. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance*, 48, 641–661.

Porteba, J., and L. Summers, 1988, "Mean Reversion in Stock Returns: Evidence and Implications," *Journal of Financial Economics*, 22, 27–60.

Richardson, M., and J. Stock, 1989, "Drawing Inferences from Statistics Based on Multi-Year Asset Returns," *Journal of Financial Economics*, 25, 323–348.

Richardson, M., and T. Smith, 1991, "Tests of Financial Models in the Presence of Overlapping Observations," *Review of Financial Studies*, 4, 227–254.

Ross, S., 1976, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360.

Rozeff, M., 1984, "Dividend Yields are Equity Risk Premiums," *Journal of Portfolio Management*, 11, 68–75.

Searle, S., 1971, *Linear Models*, Wiley, New York.

Summers, L., 1986, "Does the Stock Market Rationally Reflect Fundamental Values," *Journal of Finance*, 41, 591–601.

White, H., 1980 "A Heteroskedasticity-Consistent Covariance Matrix Estimator and Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.