THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 2 • JUNE 1993

Predictable Stock Returns: The Role of Small Sample Bias

CHARLES R. NELSON and MYUNG J. KIM*

ABSTRACT

Predictive regressions are subject to two small sample biases: the coefficient estimate is biased if the predictor is endogenous, and asymptotic standard errors in the case of overlapping periods are biased downward. Both biases work in the direction of making t-ratios too large so that standard inference may indicate predictability even if none is present. Using annual returns since 1872 and monthly returns since 1927 we estimate empirical distributions by randomizing residuals in the VAR representation of the variables. The estimated biases are large enough to affect inference in practice, and should be accounted for when studying predictability.

THE PROPOSITION THAT STOCK returns are not predictable was until very recently regarded as one of the most (some would say the only) firmly established empirical results in economics. The extent to which the nonpredictability result has been overturned in the last few years is reflected in the opening sentence of a recent paper by Fama and French (1988): "There is much evidence that stock returns are predictable." Indeed, a recent series of papers including Keim and Stambaugh (1986), Campbell and Shiller (1988), Fama and French (1988) and Cutler, Poterba, and Summers (1991) report that "fundamentals" such as dividend yield and price-earnings ratio explain 25% or more of the variation in stock returns measured over intervals of several years. Further, Balvers, Cosimano, and McDonald (1990), Schwert (1990), and Fama (1990) present evidence that economic indicators such as industrial production also have predictive power for stock returns.

This paper focuses on the possibility that small sample bias could be playing an important role in the inference that stock returns are predictable from fundamentals and in estimates of the degree of predictability. A t-ratio will be misleading if either the regression coefficient is biased or if the standard error is biased. Small sample bias in asymptotic standard errors in

* From the University of Washington and NBER, and the University of Alabama, respectively.
We are grateful to John Campbell for providing us with the data set assembled by him and Robert Shiller. Helpful comments from Stephen Cecchetti, Gregory Chow, Tim Cogley, Robert Engle, Benjamin Friedman, Robert Hodrick, Bruce Lehmann, Andrew Lo, Richard Parks, Pierre Perron, G. William Schwert, Robert Stambaugh, Richard Startz, René Stulz (the editor), Stephen Turnovsky, Kenneth West, anonymous referees, and participants in seminars at the Federal Reserve Board, the NBER, Northwestern University, Ohio State University, Princeton University, Southern Methodist University, the University of California at Santa Barbara, and the University of Oregon are acknowledged with thanks, but responsibility for any errors is entirely the authors'.

the context of overlapping observations on multiperiod returns has received attention in several recent papers which conclude that they are too small; see Hodrick (1991), Kim, Nelson, and Startz (1991), Richardson and Smith (1991), and Richardson and Stock (1989). However, the potential for small sample bias in the regression coefficient has not received corresponding attention, although it is well known that regression on lagged endogenous variables is not unbiased in finite samples. The existence of small sample bias in tests of predictability has been pointed out by Mankiw and Shapiro (1986) and Stambaugh (1986) who showed that regression on predetermined variables will reject the null hypothesis of nonpredictability too often. Nevertheless, those reporting predictability have not generally taken coefficient bias into account. Indeed, Fama and French (1988) cite Stambaugh's paper but argue that "This bias arises only when yields track time-varying expected returns. It does not bias the tests toward false conclusions that yields have forecast power."

Section I of the paper uses an approximation due to Stambaugh (1986) to show that small sample bias in the regression coefficient may be large enough to affect statistical inference in regressions where the log of dividend yield is the predictor of stock returns. Section II presents estimates of the empirical distributions of regression statistics for annual data on the S & P Index since 1872. Our methodology is based on simulation of the VAR approximation of the present value model in which artificial sequences of return and dividend yield are generated by randomization of the sample residuals. We also stratify the sample to correspond to the high variance period 1929 to 1939 for returns to explore the effect of heteroskedasticity on the sampling distributions. Results for the full sample period are also compared with those for the period since World War II during which apparent predictability was much higher. Section III studies the CRSP data set for monthly returns on NYSE stocks from 1926 which offers a much larger number of observations though for a shorter historical period. Section IV generalizes the methodology to consider predictors such as the Index of Industrial Production which are not explicit in the present value model but which are presumably endogenous in the system which also generates dividends and therefore returns. Finally, Section V offers some conclusions.

I. Small Sample Bias in the Regression of Stock Return on Dividend Yield

Table I reports regressions of real and excess returns on the lagged value of the log of the dividend yield using the annual data set of Campbell and Shiller (1988) for the Standard and Poor's Composite Index. Price is recorded for January of the years 1871 to 1987 and dividend is the amount paid during the year, so the sample period for one year total returns and lagged dividend yield is 1872 to 1986. Real and excess returns are calculated using the PPI and commercial paper rate respectively. Following Campbell and Shiller (1988), one year returns are accumulated forward over horizons of three and

Table I

Regressions of Total Return on Log of Dividend Yield S & P Annual Returns, 1872 to 1986

Least squares regressions of the total real or total excess return on the S & P Composite Index on the log of dividend yield observable. The data are annual and as described in Campbell and Shiller (1988) and were obtained from them. They deflate the S & P index and dividends by the Producer Price Index to calculate real total return and subtract from that the real return on 4 to 6 month commercial paper to obtain excess total return. In the case of overlapping three- and ten-year returns, the t-ratio is calculated using the standard error suggested by Hansen and Hodrick (1980) which recognizes the moving average structure of the regression error.

Return	Sample	Slope	t-Ratio	R^2
		One-Year Return		
Real	1872–1986	0.12	2.14	0.039
	1872–1946	0.10	1.25	0.021
	1947–1986	0.25	2.79	0.170
Excess	1872–1986	0.08	1.34	0.016
	1872–1946	0.04	0.43	0.002
	1947–1986	0.25	3.09	0.200
		Three-Year Return		
Real	1872–1984	0.35	2.46	0.110
	1872–1944	0.31	1.63	0.066
	1947–1984	0.63	2.90	0.380
Excess	1872–1984	0.30	2.07	0.079
	1872–1944	0.31	1.55	0.057
	1947–1984	0.60	3.33	0.434
		Ten-Year Return		
Real	1872–1977	0.97	3.25	0.267
	1872–1937	0.84	2.59	0.188
	1947–1977	2.02	4.43	0.714
Excess	1872–1977	0.74	2.13	0.182
	1872–1937	0.41	1.76	0.089
	1947–1977	1.95	6.36	0.794

ten years to form overlapping multiyear returns. Standard errors for the slope coefficient are adjusting for serial correlation induced by the overlap following the methodology of Hansen and Hodrick (1980), as has been standard in the empirical literature. Results for the subperiods before and after January 1947 are also reported. Kim, Nelson, and Startz (1991) report that World War II represented a sharp breaking point in the mean reversion properties of stock returns.

Key features of predictive regressions reported in the literature are apparent in Table I. In particular, the slope coefficient is positive in every case, implying that expected return is small when dividend yield is low. Further, the t-ratio and R^2 increase with return horizon, suggesting that the strongest evidence for predictability is found at long horizons. For the full period, real return is more predictable than excess return. However, when the sample is split at 1947 it becomes clear that predictability is a post–World War II phenomenon. In spite of the longer sample period there is little evidence of predictability for the prewar period. In addition, excess return has apparently been somewhat more predictable since 1947 than real return. The difference here between predictability before and after World War II stands in contrast to what has been reported in the literature on mean reversion in returns where the pattern is the reverse. Long lag autocorrelation in returns is apparent in prewar returns but seems to have disappeared after 1947; see Kim, Nelson, and Startz (1991). While the reversal in pattern may seem to present an inconsistency, Campbell (1991) shows that predictability conditional on fundamentals does not necessarily imply predictability conditional on past returns. We have also extended the data set through 1990 and find that inclusion of the additional four annual returns makes no appreciable difference in the regression results reported here (1987 was not an extraordinary year on an annual basis). We retain the 1986 ending date to preserve comparability of our results with published studies.

Taken at face value, these results lead to the inference that stock returns have been predictable from dividend yields over the past century, particularly since World War II, and that predictability increases with horizon. However, there are a number of reasons to exercise caution. As mentioned in the introduction, asymptotic standard errors in the case of overlapping returns may be downward biased in small samples, and the regression coefficient will also be biased if the predictor is jointly endogenous with the return, even if it is predetermined. While the increase in R^2 with return horizon has been emphasized in this literature as a strong and important feature of the empirical evidence, Granger and Newbold (1974) caution against interpreting high R^2 as evidence of a relationship when the data are positively autocorrelated as it is in the case of overlapping returns. Finally, the stability of regression across the pre- and postwar periods raises the issue of choosing an appropriate sample period for drawing inferences. All of these questions warrant further investigation.

That the bias in the regression coefficient may be large enough to affect inference is suggested by the work of Stambaugh (1986) and Mankiw and Shapiro (1986) who both studied the system

$$y_t = \alpha + \beta x_{t-1} + u_t; \quad u \sim \text{i.i.d.}(0, \sigma_u^2) \tag{1}$$

$$x_t = \mu + \phi x_{t-1} + v_t; \quad v \sim \text{i.i.d.}(0, \sigma_v^2) \tag{2}$$

where u and v are white noise errors that are contemporaneously correlated with covariance σ_{uv}. Although the predictive regression of y_t on x_{t-1} is consistent asymptotically, it is not unbiased in finite samples because x is not fixed in repeated samples. Stambaugh proves that the bias in the least squares estimate of β in (1) is proportional to the bias in the least squares estimate of ϕ in (2), in particular

$$E(\hat{\beta} - \beta) = (\sigma_{uv}/\sigma_v^2)E(\hat{\phi} - \phi). \tag{3}$$

To see the intuition behind this, note that in the special case that the predictor is y_{t-1} then β and ϕ are the same parameter and will have the same bias. In general, x_{t-1} may be a proxy for y_{t-1} and the factor (σ_{uv}/σ_v^2) reflects both the correlation between the innovations of the two processes and their relative scales. Kendall (1954) showed that the bias in least squares estimate of ϕ is

$$E(\hat{\phi} - \phi) = -(1 + 3\phi)/n + O(n^{-2}). \tag{4}$$

Thus, the lagged value of x may appear to be a predictor of y even though it has no predictive power. This effect will be stronger the more autocorrelated is x, the stronger is the contemporaneous relation between the innovations, and the smaller is the sample size.

The relevance of this result in the case of dividend yields predicting stock returns becomes clear from the linear approximation to the present value model (PVM) due to Campbell and Shiller (1988). They show that in the case of constant expected real return the PVM can be written as

$$r_t = \delta_t - \rho\,\delta_{t+1} + \Delta d_t + \text{const} \tag{5}$$

$$\delta_t = -\sum_{j=0}^{\infty} \rho^j E_t[\Delta d_{t+j}] + \text{const} \tag{6}$$

where r_t is total real return during year t defined as $\ln[(P_{t+1} + D_t)/P_t]$ where P is the real price at the beginning of period t and D is the real dividend paid during period t, δ_t is the log of the dividend price ratio observable at the beginning of year t, d_t is the log of the real dividend paid during year t, and ρ is a discount factor. The same representation holds whether it is real or excess return that has constant expectation. To put the model in vector ARMA form we assume that the vector of information, say z_t, used by the market to forecast dividend growth at the beginning of period t has a VAR representation, say $z_t = Az_{t-1} + w_t$, where w_t is a serially random vector of forecast errors. The list of variables in z_t includes Δd_t as the first element, but we do not need to be specific about the other variables. Substituting for the forecasted future Δd_t in the equation for d_t one obtains

$$r_t = \alpha + e'[l - \rho A]^{-1} w_{t+1} \tag{7}$$

$$\delta_{t+1} = c - e'[l - \rho A]^{-1} A[l - AL]^{-1} w_{t+1} \tag{8}$$

where e' is the vector $(1, 0, \dots)$ which selects the first element of a column vector, c is a constant which is a function of the parameters, and L is the lag operator. Note that the return during period t is a linear combination of only new information w_{t+1} revealed during t. Further, the innovation in the MA process for δ is $e'[l - \rho A]^{-1} A w_{t+1}$ which is also a linear combination of the innovations in the information process but not the same combination as the return. Thus the innovations in the processes for r_t and δ_{t+1} will be contemporaneously correlated but not at leads or lags.

In the case that expected excess return is constant, r_t represents the difference between the total return on stocks and the return on an alternative asset and Campbell and Shiller show that Δd_{t+j} is replaced in (5) and (6) by $[\Delta d_{t+j} - ra_{t+j}]$ where ra_t is return on the alternative asset during period t. The rest of the analysis follows with $[\Delta d_{t+j} - ra_{t+j}]$ playing the role of Δd_{t+j} leading to a representation of the form of (7) and (8).

If the dividend yield series δ_t is AR1 then the linearized PVM corresponds to the bivariate process studied by Stambaugh and by Mankiw and Shapiro. For the annual S & P data 1872 to 1986 an AR1 specification for δ_t is a reasonable one with $\hat{\phi} = 0.71$ and the sample value of (σ_{uv}/σ_v^2) is -0.68 for real returns and -0.72 for excess. The implied coefficient bias in the regression of one year returns on δ_t is 0.02. Multiperiod returns correspond to observing the dividend yield process at three- and ten-year intervals. If the observations were nonoverlapping the implied coefficient bias for three-year returns would be 0.05 and for ten-year returns 0.11. Clearly these are not large enough to explain away the point estimates in Table I, but they are sufficient to influence statistical inference. For one-year real returns, reducing the coefficient by 0.02 reduces the t-statistic from 2.14 to 1.78. The potential to alter inferences is indicated by the fact that none of the excess return regressions for the full sample period would remain significant at the 0.05 level if such an adjustment is appropriate.

The next section of the paper studies the statistical issues suggested by this discussion using simulation based on the VAR representation of return and yield. In particular, we are interested in the small sample distribution of the coefficient bias in the case of overlapping observations, the properties of asymptotic standard errors and t-ratios, the distribution of R^2, and the effect of the high variance period 1929 to 1939 on these sampling distributions.

II. Empirical Sampling Distributions for Regression Statistics: Annual Returns on the S & P Composite, 1872 to 1986.

A number of strategies exist for estimating sampling distributions. Monte Carlo requires that we assume a distribution function for the disturbances, usually the Normal. In the case of stock returns the distribution is known to be non-Normal and heteroskedastic. In this paper we simulate artificial histories of the pair (r_t, δ_t) using the fitted VAR approximation to the present value model and random drawings of the residual pair $(\hat{u}_t, \hat{v}_t)$. The initial observation on δ_t is drawn from a normal distribution with mean equal to the historical mean and variance $\sigma_v^2/(1 - \phi^2)$ implied by its AR(1) representation. The resulting artificial return and yield data are consistent with the present value model under the restriction that returns are not predictable, but will also have serial correlation and dispersion similar to the historical time series; see Cogley (1991). We draw residual pairs without replacement which is called randomization; see Noreen (1989) for extensive discussion. Randomization differs from bootstrapping only in that sampling is without replacement and it is an attractive strategy in situations where the population distribution is unknown and the null hypothesis involves absence of a relationship. In the first experiment reported here all the residual pairs are put in a single urn to generate the full sample under homoskedasticity. Following Kim, Nelson, and Startz (1991), we then stratify the sample in a second experiment, placing the residuals from the high variance period 1929 to 1939 in a separate urn and then drawing from the low or high variance urn according to position in the sequence. This stratified randomization gives us information about the sensitivity of sampling distribution to the particular pattern of heteroskedasticity that occurred historically; see Schwert (1989). In each case the simulations were repeated 1000 times. The sampling distributions are summarized in the tables which follow by their mean, median if the mean is sometimes undetermined, fractiles, and the estimated probability of obtaining a statistic at least as large as the historical value. The latter are denoted "One tail p-value" and allow readers to draw their own conclusions in regard to statistical significance.

Table II summarizes the estimated sampling distribution for the slope coefficient, R^2, and t-ratios in regressions of annual total returns for the S & P Composite Index on the lagged log of dividend yield for the period 1872 to 1986. The bias in the slope coefficient is consistent with Stambaugh's approximation at all horizons if one takes the relevant number of observations to be the number of possible nonoverlapping observations rather than the number of actual overlapping observations used in the regression.

The sampling distribution of R^2 is tabulated in the second panel of Table II. Upward bias in R^2 for regressions of returns on fundamentals has been previously noted by Kandel and Stambaugh (1989). In spite of bias, the sample values of R^2 are seen to be large relative to their sampling mean. It is important to note, however, that the upper tail of the distribution of R^2 does not correspond to large deviations of the t-ratio from its mean because the t-ratio is not centered around zero due to bias in the slope coefficient. Thus, the usual correspondence between large values of R^2 and the rejection region under the null hypotheses of no relationship breaks down in this situation.

In the case of overlapping multiperiod returns authors have generally adopted the method of Hansen and Hodrick (1980), hereafter HH, which adjusts standard error for induced autocorrelation recognizing that if one period returns are serially random then the implied structure of the regression error is $MA(K-1)$ where K is the return horizon. We denote the resulting t-ratio by HHt1. Positive definiteness of the covariance matrix is not guaranteed so the HH standard error may not exist for a given sample. Further, the returns data show clear heteroskedasticity which should be taken into account in standard errors. Newey and West (1987), hereafter NW, study a class of autocorrelation and heteroskedasticity consistent variance estimators which are positive semidefinite. The key to positive definiteness is the weighting of the sample autocorrelations and they show that the scheme

Table II

Randomization Estimate of Distributions of Slope, R^2, and t-Ratio: S & P Annual Returns, 1872 to 1986

Least squares regressions of the real or excess total return for the S & P Composite Index on the log of dividend yield. The data are annual and as described in Campbell and Shiller (1988) and were obtained from them. They deflate the S & P index and dividends by the Producer Price Index to calculate real total return and subtract from that the real return on 4 to 6 month commerical paper to obtain excess total return. OLSt denotes ordinary least squares t-ratio. In the case of overlapping three- and ten-year returns, t-ratios are calculated using alternative standard errors designed to take into account serial correlation in the regression error: that suggested by Hansen and Hodrick (1980) denoted HHt1, the positive semidefinite and heteroskedasticity-consistent standard error of Newey and West (1987) with lag alternatively set at the data overlap less one, denoted NWt1, and twice that lag, denoted NWt2, and finally HHt2, which is a heteroskedasticity-consistent version of HH. Mean, median, and fractiles of the sampling distributions and the one tail p-value for the historical statistic are estimated from 1000 artificial histories of return and log of dividend yield generated by randomization of historical residuals from the estimated VAR approximation to the present value model under the null hypothesis of constant expected return as described in Section I.

	Slope Coefficient				
Return	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
			One-Year Return		
Real	0.12	0.02	−0.09	0.14	0.040
Excess	0.08	0.02	−0.09	0.15	0.158
			Three-Year Return		
Real	0.35	0.05	−0.23	0.36	0.027
Excess	0.30	0.05	−0.23	0.38	0.052
			Ten-Year Return		
Real	0.97	0.13	−0.59	0.80	0.006
Excess	0.74	0.14	−0.61	0.78	0.040

	R^2			
Return	Historical	Mean	0.95 Fractile	One Tail p-Value
		One-Year Return		
Real	0.039	0.009	0.032	0.032
Excess	0.016	0.009	0.035	0.181
		Three-Year Return		
Real	0.110	0.021	0.079	0.020
Excess	0.079	0.022	0.077	0.049
		Ten-Year Return		
Real	0.267	0.042	0.159	0.004
Excess	0.182	0.043	0.157	0.035

Table II—*Continued*

Return	Type	t-Ratio Historical	Mean	Median	0.025 Fractile	0.975 Fractile	One Tail p-Value
			One-Year Return				
Real	OLSt	2.14	0.22	0.24	−1.76	2.13	0.024
Excess	OLSt	1.34	0.24	0.25	−1.81	2.21	0.131
			Three-Year Return				
Real	HHt1	2.46	0.29	0.30	−1.74	2.52	0.027
	HHt2	2.75	0.35	0.31	−1.94	2.84	0.029
	NWt1	3.09	0.37	0.35	−2.17	3.03	0.025
	NWt2	2.83	0.37	0.34	−2.09	2.98	0.036
Excess	HHt1	2.07	0.30	0.28	−1.72	2.52	0.051
	HHt2	2.26	0.35	0.31	−1.74	2.98	0.055
	NWt1	2.52	0.37	0.35	−2.00	3.18	0.051
	NWt2	2.34	0.37	0.34	−1.98	3.17	0.061
			Ten-Year Return				
Real	HHt1	3.25	0.50	0.41	−1.60	3.02	0.018
	HHt2	3.51	*	0.57	−2.22	10.31	0.089
	NWt1	3.41	0.62	0.53	−2.05	3.83	0.042
	NWt2	4.25	0.69	0.57	−2.28	4.25	0.026
Excess	HHt1	2.13	0.55	0.45	−1.47	3.11	0.095
	HHt2	3.10	**	0.56	**	5.59	0.103
	NWt1	3.35	0.71	0.59	−2.17	4.18	0.063
	NWt2	3.26	0.78	0.63	−2.26	4.60	0.081

* Negative variance in 29 out of 1000 samples.

** Negative variance in 27 out of 1000 samples and consequently estimate of 0.025 fractile not available.

where lag j is given weight $w(j, m) = 1 - [j/m + 1]$, where m is the maximum lag, has this property. Consistency requires that m increase with sample size, so m is not to be identified with the order of the moving average. As Newey and West point out, "The specification of an appropriate growth rate for $m(T)$ gives little guidance concerning the choice of m in practice." Since the weights are less than one and decline with lag it has been suggested setting m larger than the known order of the MA may be attractive; Cochrane (1991) suggests using $2(K - 1)$. We experiment with $m = (K - 1)$ and $m = 2(K - 1)$. The resulting t-ratios are denoted here by NWt1 and NWt2 respectively. Setting all the weights equal to one and m equal to $(k - 1)$ gives a heteroskedastic-consistent version of HH which we call HHt2.

The sampling distribution seen in the third panel of Table II suggests the following generalizations: The t-ratios are biased upward by an amount that increases with return horizon. For overlapping returns, the spread of the distribution is also too large. In the case of three-year returns the distance between the 0.025 and 0.975 fractiles is about 4.5 for HHt1, 4.75 for HHt2,

Table III

Stratified Randomization Distribution of Slope, R^2, and t-Ratio: S & P Annual Returns, 1872 to 1986

Least squares regressions of the real or excess total return for the S & P Composite Index on the log of dividend yield. The data are annual and as described in Campbell and Shiller (1988) and were obtained from them. They deflate the S & P index and dividends by the Producer Price Index to calculate real total return and subtract from that the real return on 4 to 6 month commercial paper to obtain excess total return. OLSt denotes ordinary least squares t-ratio. In the case of overlapping three- and ten-year returns, t-ratios are calculated using alternative standard errors designed to take into account serial correlation in the regression error: that suggested by Hansen and Hodrick (1980) denoted HHt1, the positive semidefinite and heteroskedasticity-consistent standard error of Newey and West (1987) with lag alternatively set at the data overlap less one, denoted NWt1, and twice that lag, denoted NWt2, and finally HHt2, which is a heteroskedasticity-consistent version of HH. Mean, median, and fractiles of the sampling distributions and the one tail p-value for the historical statistic are estimated from 1000 artificial histories of return and log of dividend yield generated by stratified randomization of historical residuals from the estimated VAR approximation to the present value model under the null hypothesis of constant expected return as described in Section I in which residuals from the high variance period 1929 to 1939 are treated as a separate population.

	Slope Coefficient				
Return	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
		One-Year Return			
Real	0.12	0.04	−0.07	0.17	0.104
Excess	0.08	0.04	−0.07	0.19	0.264
		Three-Year Return			
Real	0.35	0.11	−0.16	0.43	0.065
Excess	0.30	0.11	−0.19	0.42	0.218
		Ten-Year Return			
Real	0.97	0.24	−0.53	0.81	0.009
Excess	0.74	0.22	−0.41	0.77	0.040
		R^2			
			0.95 Fractile		
		One-Year Return			
Real	0.039	0.011	0.042		0.061
Excess	0.016	0.012	0.048		0.240
		Three-Year Return			
Real	0.110	0.027	0.098		0.037
Excess	0.079	0.027	0.098		0.088
		Ten-Year Return			
Real	0.267	0.051	0.173		0.008
Excess	0.182	0.047	0.175		0.004

Table III—*Continued*

		t-Ratio					
Return	Type	Historical	Mean	Median	0.025 Fractile	0.975 Fractile	One Tail p-Value
		One-Year Return					
Real	OLSt	2.14	0.59	0.56	−1.31	2.48	0.058
Excess	OLSt	1.34	0.63	0.60	−1.21	2.61	0.226
		Three-Year Return					
Real	HHt1	2.46	0.74	0.67	−1.13	2.87	0.054
	HHt2	2.75	0.82	0.72	−1.23	3.18	0.052
	NWt1	3.09	0.87	0.80	−1.38	3.36	0.042
	NWt2	2.83	0.87	0.78	−1.33	3.34	0.065
Excess	HHt1	2.07	0.73	0.68	−1.24	2.86	0.111
	HHt2	2.26	0.79	0.70	−1.33	3.20	0.098
	NWt1	2.52	0.83	0.76	−1.45	3.29	0.083
	NWt2	2.34	0.84	0.76	−1.45	3.26	0.107
		Ten-Year Return					
Real	HHt1	3.25	0.90	0.86	−1.35	3.36	0.030
	HHt2	3.51	*	1.11	−1.46	*	0.158
	NWt1	3.41	1.15	1.03	−1.62	4.56	0.082
	NWt2	4.25	1.27	1.13	−1.60	5.06	0.055
Excess	HHt1	2.13	0.84	0.77	−1.22	3.15	0.128
	HHt2	3.10	**	0.98	−1.63	23.85	0.161
	NWt1	3.35	1.08	0.93	−1.51	4.18	0.071
	NWt2	3.26	1.19	0.97	−1.65	4.75	0.114

* Negative variance in 29 out of 1000 samples, therefore 0.975 fractile not available.
** Negative variance in 27 out of 1000 samples.

and 5 for both versions of NW compared with about 4 for a t distribution. At the ten-year horizon the spread has widened to about 4.6 for HHt1, about 6 for NWt1, and more than 6 for NWt2. The spread for HHt2 is not reliably estimated because the variance failed to be positive in about 3% of the runs. Perhaps surprisingly, this problem was not encountered in the case of HHt1. The original HH t-ratio seems to have a less distorted sampling distribution than the heteroskedasticity-consistent versions, however these artificial histories are homoskedastic. Doubling the lag length used in calculating NW makes no practical difference except that at horizon ten years the longer lag length has larger bias and greater dispersion. The combined effect of the two biases in the t-ratio is that the tabulated p-values for the historical regressions are much larger than would be implied if one assumed $t(115)$ to be appropriate distributions. Correspondingly, correct upper tail critical values are much larger than under $t(115)$. For example, to do a two tail test at the 0.05 level of significance using ten-year returns the critical value is about 3 for HHt1 and about 4 for NW instead of about 2.

Table IV

Randomization Distributions of Slope Coefficient, R^2, and t-Ratio: Postwar S & P Annual Returns, 1947 to 1986

Least squares regressions of the real or excess total return for the S & P Composite Index on the log of dividend yield. The data are annual and as described in Campbell and Shiller (1988) and were obtained from them. They deflate the S & P index and dividends by the Producer Price Index to calculate real total return and subtract from that the real return on 4 to 6-month commercial paper to obtain excess total return. OLSt denotes ordinary least squares t-ratio. In the case of overlapping three- and ten-year returns, t-ratios are calculated using the standard error suggested by Hansen and Hodrick (1980), denoted HHt1, designed to take into account serial correlation in the regression error. Mean, median, and fractiles of the sampling distributions and the one tail p-value for the historical statistic are estimated from 1000 artificial histories of return and log of dividend yield generated by randomization of historical residuals from the estimated VAR approximation to the present value model under the null hypothesis of constant expected return as described in Section I.

	Slope Coefficient				
Return	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
			One-Year Return		
Real	0.25	0.09	−0.12	0.41	0.116
Excess	0.25	0.08	−0.11	0.37	0.088
			Three-Year Return		
Real	0.63	0.23	−0.30	0.87	0.190
Excess	0.60	0.19	−0.31	0.78	0.094
			Ten-Year Return		
Real	2.02	0.59	−0.80	1.77	0.007
Excess	1.95	0.50	−0.78	1.62	0.003

	R^2			
Return	Historical	Mean	0.95 Fractile	One Tail p-Value
		One-Year Return		
Real	0.170	0.033	0.126	0.019
Excess	0.200	0.032	0.123	0.006
		Three-Year Return		
Real	0.380	0.079	0.294	0.016
Excess	0.434	0.077	0.283	0.006
		Ten-Year Return		
Real	0.714	0.197	0.649	0.011
Excess	0.794	0.182	0.550	0.003

Table IV—*Continued*

		t-Ratio					
Return	Type	Historical	Mean	Median	0.025 Fractile	0.975 Fractile	One Tail *p*-Value
		One-Year Return					
Real	OLSt	2.79	0.60	0.57	−1.28	2.64	0.015
Excess	OLSt	3.09	0.56	0.57	−1.34	2.61	0.006
		Three-Year Return					
Real	HHt1	2.90	0.79	0.73	−1.35	3.45	0.050
Excess	HHt1	3.33	0.75	0.66	−1.48	3.27	0.025
		Ten-Year Return					
Real	HHt1	4.43	1.44	1.19	−1.39	5.61	0.050
Excess	HHt1	6.36	1.28	1.11	−1.38	4.70	0.009

To see what effect heteroskedasticity has on the sampling distributions, the annual data have been stratified according to whether an observation falls within the high variance period 1929 to 1939 or not, as described above. In Table III, we see that the resulting heteroskedasticity in the data roughly doubles the positive bias in the slope coefficient and *t*-ratios, and this is reflected in higher values of R^2. The dispersion of the *t*-ratios does not change substantially. Surprisingly the NW versions, which are heteroskedasticity consistent, do not perform better than the original HHt1. Again, doubling the number of lags used in NW has little effect except at horizon ten years and in that case the smaller number is preferred. Taking into account the effect of heteroskedasticity on sampling distributions clearly weakens the evidence for predictability, and that which remains is stronger for real returns than for excess, and at longer horizons.

It was clear from Table I that returns have been more predictable since World War II, particularly so for excess returns. Fama and French (1988) reported similar results for annual real returns on the CRSP value-weighted NYSE protfolio for sample periods 1941 to 1986 and 1957 to 1986. Should World War II be treated as a breaking point and inference carried out separately for the postwar period? A Chow test on the equality of parameters in pre- and postwar one-year return regressions gives mixed results: the significance level is 0.08 for excess returns but real returns are consistent with unchanged coefficients. It should be noted that the significance level in the case of excess returns is sensitive to the specific year chosen as the breaking point. Rather than judging this issue, we have done a parallel analysis of the 1947 to 1986 returns and present the results in Table IV. We do not stratify the postwar sample.

Comparing Table IV with Table II we see that the bias in the slope coefficient is about four times as large under homoskedasticity for the postwar period as for the full period. The bias is larger because of the smaller sample size and also because of greater persistence of the log dividend yield process during the postwar period (the AR coefficient is 0.83 compared with 0.71) and a stronger relation between the innovations $((\sigma_{uv}/\sigma_v^2) = -0.98$ for real returns and -0.78 for excess returns). The bias is also about twice as large as for the stratified full sample. Since the point estimates of the slope are also considerably larger for the postwar period, the changes in p-values from Table II to Table IV are mixed. The t-ratios reported in Table IV for overlapping returns include only the HHt1 formulation to save space, since it is the HHt1 version that departs least from the t-distribution. Although the small sample bias in the t-ratios is also much larger for this shorter period, the p-values imply stronger evidence for predictability in the postwar period, particularly for excess returns.

The simulations take as given the AR coefficient for the dividend yield process, ϕ. However, our simulations use the least squares estimate, $\hat{\phi}$, which is known to be biased downward. Since the estimated bias in the slope coefficient, $\hat{\beta}$, varies directly with the assumed value of ϕ, our procedure provides conservative estimates of the bias in the regression slope. For the relatively small number of 40 postwar observations the bias in $\hat{\phi}$ is roughly -0.1. Using the augmented Dickey-Fuller test we find that the point estimate is not significantly different from one (the t-ratio for the unit root hypothesis is only -1.84); see Dickey and Fuller (1979). Taking the unit value for ϕ as a worst case possibility for underestimation of bias, we have repeated the randomization of the postwar data and find that the means of the t-ratios roughly double; for example, the mean of OLSt for one-year returns increases from 0.60 to 1.4 and the historical t-ratio is no longer significant. This confirms that we have been conservative in estimating the bias towards finding predictability.

III. Predictability of Monthly NYSE Returns, 1926 to 1986

Fama and French (1988) reported regressions of monthly real returns for value-weighted and equally weighted portfolios based on the CRSP files for NYSE stocks 1926 to 1986 on dividend yield. This is a particularly interesting data set for our purposes because we would like to see whether more frequent observation of returns, resulting in a much larger number of observations, is sufficient to render small sample bias negligible. For one-month returns the only issue is coefficient bias since regression errors are serially uncorrelated under the null hypothesis. Since Fama and French reported only the results for real returns, noting that results for excess returns are very similar, we follow suit.

Table V presents historical regressions and estimates of the sampling distributions for the full sample period under homoskedasticity and under heteroskedasticity using stratification as with the annual data. For each regression statistic the unstratified sampling distribution is reported in the first row and the stratified in the second row. We find that bias is not small relative to historical point estimates, and stratification results in roughly a doubling of the estimated bias. As a point of reference, a t-statistic needs to be greater than about 3.1 rather than 2 to be significant at the 0.05 level, based on stratified sampling. The estimated p-values for the historical slopes and t-ratios do not represent strong evidence of predictable returns.

Note that the bias estimates for the t-ratio reported in Table V are comparable to those of Tables II and III for annual data. (The slope coefficients are not directly comparable because we follow Fama and French in not taking the log of dividend yield.) It may seem surprising that the very much larger number of monthly observations (roughly seven times as many) does not result in greatly reduced bias. Recalling Stambaugh's approximation, the

Table V

Randomization Distributions of Slope, R^2, and t-Ratio: Monthly NYSE Real Returns, 1927 to 1986

Least squares regressions of real total return for value-weighted and equally weighted portfolios of NYSE stocks on dividend yield using data from the CRSP files. The data are monthly and total return is adjusted by the Consumer Price Index to obtain real return. Dividend yield is the running total of prior 12 months. Mean and fractiles of the sampling distributions and the one tail p-value for the historical statistic are estimated from 1000 artificial histories of return and dividend yield generated by randomization of historical residuals from the estimated VAR approximation to the present value model under the null hypothesis of constant expected return described in Section III. In a companion experiment the residuals are stratified according to whether they belong to the high variance period 1929 to 1939 and those are treated as a separate population, 1927 to 1986.

Return	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
			Slope Coefficient		
Value-weighted	0.27	0.09	−0.19	0.45	0.14
stratified		0.19	−0.21	0.72	0.33
Equally weighted	0.18	0.12	−0.26	0.67	0.34
stratified		0.31	−0.22	0.96	0.64
			t-Ratio		
Value-weighted	1.70	0.40	−1.46	2.21	0.09
stratified		0.87	−1.47	3.08	0.23
Equally weighted	0.92	0.43	−1.53	2.40	0.30
stratified		1.14	−1.24	3.16	0.40
			R^2		
			0.95 Fractile		
Value-weighted	0.004	0.001	0.005		0.11
stratified		0.003	0.010		0.25
Equally weighted	0.001	0.002	0.006		0.42
stratified		0.004	0.012		0.67

bias will diminish with sample size, but it also depends on the AR coefficient of the predictor as well as on the relation between the innovations. Since monthly observations on the dividend yield process are much smoother than annual observations they have a larger AR coefficient ($\hat{\phi}$ is 0.97 for both value- and equally weighted portfolios) and the innovations are also more strongly cross-correlated in the monthly data. When observations are made at finer intervals the effects of larger sample size and greater smoothness tend to be offsetting, so bias depends primarily on the length of the time period covered by the sample. This point has been made by Perron (1989) and Pierse and Snell (1992) in the context of testing for a unit root. Intuitively, inference about β is directly related to inference about the degree of stationarity in the predictor series and that will be little influenced by having more frequent observations but will be influenced by having a longer historical record.

When the sample period is restricted to the postwar years 1947 to 1986 in Table VI, both slope coefficients and t-ratios are substantially larger as in the case of the annual data. Evidence of predictability is quite strong for value-weighted returns. In spite of upward bias, the t-ratio of 2.50 corresponds to a one tail p-value of 0.03.

Table VI

Randomization Distributions of Slope, R^2, and t-Ratio: Postwar Monthly NYSE Real Returns, 1947 to 1986

Least squares regressions of real total return for value-weighted and equally weighted portfolios of NYSE stocks on dividend yield using data from the CRSP files. The data are monthly and total return is adjusted by the Consumer Price Index to obtain real return. Dividend yield is the running total of prior 12 months. Mean and fractiles of the sampling distributions and the one tail p-value for the historical statistic are estimated from 1000 artificial histories of return and dividend yield generated by randomization of historical residuals from the estimated VAR approximation to the present value model under the null hypothesis of constant expected return described in Section III. Postwar period 1947 to 1986.

Return	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
			Slope Coefficient		
Value-weighted	0.41	0.18	−0.17	0.77	0.16
Equally weighted	0.28	0.22	−0.19	0.86	0.33
			t-Ratio		
Value-weighted	2.50	0.66	−1.19	2.60	0.03
Equally weighted	1.59	0.72	−1.17	2.60	0.20
			R^2		
			0.95 Fractile		
Value-weighted	0.013	0.003	0.011		0.03
Equally weighted	0.005	0.003	0.011		0.22

IV. Macroeconomic Variables as Predictors of Stock Returns

The presence of small sample bias is not limited to regressions of return on financial variables such as dividends or earnings which are directly related to share valuation. Regression of return on the lagged value of any variable that is endogenous to the system which determines return will in general be biased, even if the true conditional expectation of return does not depend on the lagged predictor. The assumptions of the Gauss-Markov theorem are not met in such cases; the OLS coefficient is a ratio of random variables and its expected value is not in general equal to the true slope coefficient. Macroeconomic variables are presumably determined jointly with stock returns since shocks such as innovations in monetary policy or oil prices will affect both. While we may not know the details of the system in which return and other variables are determined, we can estimate the vector process which summarizes the joint distribution of return and its prospective predictor and simulate the empirical distribution of regression statistics.

Economic activity has been reported to be a predictor of stock returns in three recent papers cited in the Introduction. For example, Balvers, Cosimano, and McDonald (1990) reported that the log of the Index of Industrial Production (IP) along with a time trend explain about 20% of the variation in annual one-year real returns during the period 1947 to 1987. The slope coefficient for IP is negative and has a t-ratio of greater than 3 in absolute value (see their Tables I and II). To estimate the bias in this regression we look for a VAR representation of stock returns and IP which can be used to summarize their joint distribution.

Our data set consists of the annual returns on the S & P used above and IP for November since, following Balvers *et al.*, that number is known to market participants at the end of the year. We find that IP exhibits little serial correlation in first differences during this period and that the unit root hypothesis is accepted using a Dickey-Fuller test. We therefore model IP as a random walk with drift. The log first differences of IP are only weakly correlated with the annual return on the S & P Index in the same year, but the correlation between return and the growth rate of IP in the following year is +0.55. Under the restriction that IP does not predict stock returns, a simple VAR representation which captures the leading indicator property of stock returns is

$$\mathrm{IP}_t = \mathrm{IP}_{t-1} + 0.037 + z_t; \quad \sigma_z^2 = 0.004 \tag{9}$$

$$r_t = 0.08 + u_t; \quad \sigma_u^2 = 0.026 \tag{10}$$

where the covariance between u_t and z_{t+1} is 0.0055. This representation corresponds to equations (1) and (2) with IP_{t+1} playing the role of x_t and z_{t+1} the role of v_t. The predictor is then x_{t-2} rather than x_{t-1}.

Table VII presents results for the predictive regression

$$r_t = \alpha + \beta\,\mathrm{IP}_{t-1}(+\gamma\,\mathrm{time}) \tag{11}$$

Table VII

Regression of S & P Annual Real Return on Industrial Production, 1948 to 1986

Least squares regressions of annual real returns for the S & P Composite Index on the Index of Industrial Production for the prior year, alternatively without and with time included. Annual returns are from Campbell and Shiller (1988) and Industrial Production is for November of the prior year. Mean, fractiles, and one tail p-value for the historical statistics are estimated from a sample of 1000 regressions using artificial histories generated by randomizing the VAR representation described in Section IV under the null hypothesis that returns are not predictable. The t-ratio reported for one-year returns uses the OLS formula. In the case of three- and ten-year returns the t-ratio uses the standard error of Hansen and Hodrick (1980).

	Slope Coefficient				
Time Trend	Historical	Mean	0.025 Fractile	0.975 Fractile	One Tail p-Value
		One-Year Return			
Without time	−0.094	−0.01	−0.14	0.11	0.11
With time	−1.14	−0.28	−0.90	0.25	0.003
		Three-Year Return			
Without time	−0.312	−0.03	−0.41	0.36	0.08
With time	−1.66	−0.71	−2.28	0.83	0.11
		Ten-Year Return			
Without time	−1.38	−0.03	−1.44	1.60	0.03
With time	−3.10	−1.16	−4.26	2.26	0.13
			t-Ratio		
		One-Year Return			
Without time	−1.62	−0.24	−2.26	1.70	0.08
With time	−3.64	−0.99	−3.08	0.89	0.003
		Three-Year Return			
Without time	−2.04	−0.26	−2.86	2.04	0.08
With time	−2.60	−1.26	−4.56	1.22	0.14
		Ten-Year Return			
Without time	−3.67	−0.52	−8.74	5.70	0.12
With time	−1.90	−1.54	−5.75	2.06	0.38

where we alternatively include time as a regressor or leave it out to see the effect that detrending has on bias and inference. The historical point estimates of β are all negative and larger in absolute value when time is included. We estimated the empirical distribution of these statistics from artificial histories of r and IP generated by randomizing residuals in the constrained VAR representation. Note that the coefficient bias is also larger in absolute value when time is included. The negative sign of the bias is predicted by Stambaugh's approximation, since the VAR innovations are positively correlated in this case (the additional lag in the predictor does not change the sign).

The fact that the bias is smaller in absolute value when time is excluded is related to the bias in estimating the unit root of the IP process. Evans and Savin (1984) showed that the downward bias in the least squares estimate of the unit root diminishes with the ratio of the drift to the standard deviation of the innovation, and that ratio is large enough in the case of IP to reduce the bias in $\hat{\phi}$ and therefore, given (1), the bias in $\hat{\beta}$. If IP were a driftless random walk the bias in $\hat{\beta}$ would be about -0.12 for one-year returns, but the substantial upward drift in IP relative to its standard deviation has the effect of shrinking this bias to -0.01 as seen in Table VII. In contrast, including time in the regression is seen to magnify the bias in $\hat{\beta}$ to -0.24. It is straightforward to show that the bias in this regression is again given by (1) but where $\hat{\phi}$ is interpreted as the least squares estimate of the AR coefficient for the detrended predictor and ϕ as the AR coefficient in the original predictor series. Nelson and Kang (1981) show that $E(\hat{\phi} - \phi)$ in the random walk case is approximately $-10/n$ where n is sample size. This accounts for the empirical bias of -0.28 reported in Table VII.

Only the HHt1 t-ratios are reported for overlapping returns because that was again the version which departed least from a t-distribution. The strongest evidence for predictability comes from the one-year return regression with time included, where the t-ratio for IP is -3.64. In spite of an estimated bias of -0.99, this statistic remains in the extreme tail of the sampling distribution with a one tail p-value of 0.003. Going to three- and ten-year returns the bias increases due to smaller effective sample size, and so does the spread of the sampling distribution since estimated standard errors are again too small. Reflecting the combined effect of these two biases 0.025 fractile falls at -4.56 for three-year returns and -5.75 for ten-year returns. Thus the t-ratio of -2.60 for predicting three-year returns has a p-value of 0.14.

V. Summary and Conclusions

The t-ratios from predictive regressions of stock returns on the lagged values of financial fundamentals or macroeconomic indicators are subject to two small sample biases that both work in the direction of indicating that returns are more predictable than they in fact are. First, the slope coefficient is biased if the predictor is endogenous in the system that generates returns, even if it is predetermined. Stambaugh (1986) showed that the bias is a decreasing function of the sample size and an increasing function of the autocorrelation in the predictor and of the contemporaneous correlation between innovations in the two variables. Second, in the case of overlapping multiperiod returns, standard errors based on asymptotic theory tend to be too small in finite samples.

Our methodology for studying these biases is to model the variables as a VAR under the null hypothesis, and then generate artificial histories of them using the estimated VAR and randomized sequences of the historical residuals. The sequence is also stratified according to periods of high and low variance in order to simulate the historical pattern of heteroskedasticity. Repeated regressions for the artificial histories give us an empirical sampling distribution. We investigate four versions of standard errors designed to account for serial correlation in regression errors introduced by overlapping of multiperiod returns, three of which accommodate heteroskedasticity.

The main finding of the paper is that both sources of small sample bias are important, and are large enough to mitigate evidence that the lagged value of the dividend yield is a predictor of stock returns. Using annual returns for 1872 to 1986 we find that one tail p-values estimated from the empirical distributions are substantially larger than what would be implied if the t-distribution were appropriate. Stratification of the sample suggests larger biases and less significant results. Perhaps surprisingly, heteroskedastic consistent versions of the standard error do not perform better than simpler versions. Indeed for all four versions, the spread of the distribution of the t-ratio is much in excess of that of the t-distribution, suggesting that simulation is essential to proper assignment of significance levels.

When the sample is shortened to include only the period since 1947 it becomes apparent that predictability of returns from fundamentals has been primarily a phenomenon of the post–World War II era. If it is appropriate to treat this as a separate period, then the observed predictability is highly significant in spite of larger small sample bias associated with a smaller sample size.

Monthly returns data offers far larger sample size than annual data, but, perhaps surprisingly, the regression bias is not mitigated by this. The important factor in determining small sample bias is not the number of observations per se, but the length of the historical record.

Finally, predicting returns using detrended industrial production provides a more general example of regression on lagged values of macroeconomic variables. Modeling the variables in a VAR in which return is not predictable, we find that detrending substantially increases the bias. Nevertheless, regressions predicting one-year returns remain highly significant although there is not strong evidence of predictability at longer horizons.

The conclusion which we draw from these experiments is that valid inferences cannot be drawn from predictive regressions using conventional tables that are appropriate in the case of classical regression. The investigator would seem to be obliged to develop the empirical distribution of the statistic under the null hypotheses using simulations methods before drawing inferences.

REFERENCES

Balvers, Ronald J., Thomas F. Cosimano, and Bill McDonald, 1990, Predicting stock returns in an efficient market, *Journal of Finance* 45, 1109–1128.

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

——— and Robert J. Shiller, 1988, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Cochrane, John H., 1991, Production-based asset pricing and the link between stock returns and economic fluctuations, *Journal of Finance* 46, 209–237.

Cogley, Timothy, 1991, Out-of-sample predictive power of dividend yield regressions, Unpublished manuscript, University of Washington.

Cutler, David M., James M. Poterba, and Lawrence H. Summers, 1991, Speculative dynamics, *Review of Economic Studies* 58, 529–546.

Dickey, David A., and Wayne A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association* 74, 427–431.

Evans, G. B. A., and N. E. Savin, 1984, Testing for unit roots: 2, *Econometrica* 52, 1241–1269.

Fama, Eugene F., 1990, Stock returns, real returns, and economic activity, *Journal of Finance* 45, 1089–1108.

——— and Kenneth R. French, 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3–25.

Granger, C. W. J., and P. Newbold, 1974, Spurious regressions in econometrics, *Journal of Econometrics* 2, 111–120.

Hansen, Lars Peter, and Robert J. Hodrick, 1980, Forward exchange rates as optimal predictors of future exchange rates: An econometric analysis, *Journal of Political Economy* 88, 829–853.

Hodrick, Robert J., 1991, Dividend yields and expected stock returns: Alternative procedures for inference and measurement, Technical Working Paper No. 108, National Bureau of Economic Research.

Kandel, Shmuel, and Robert F. Stambaugh, 1989, Modeling expected stock returns for long and short horizons, Unpublished manuscript, University of Pennsylvania.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357–390.

Kendall, M. G., 1954, Note on bias in the estimation of autocorrelation, *Biometrika* 41, 403–404.

Kim, Myung-Jig, Charles R. Nelson, and Richard Startz, 1991, Mean reversion in stock prices: A reappraisal of the empirical evidence, *Review of Economic Studies* 58, 515–528.

Mankiw, N. Gregory, and Matthew D. Shapiro, 1986, Do we reject too often?, *Economics Letters* 20, 139–145.

Nelson, Charles R., and Heejoon Kang, 1981, Spurious periodicity in inappropriately detrended time series, *Econometrica* 49, 741–751.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Noreen, Eric, 1989, *Computer-Intensive Methods for Testing Hypotheses: An Introduction* (John Wiley and Sons, New York).

Perron, Pierre, 1989, Testing for a random walk: A simulation experiment of power when the sampling interval is varied, in B. Raj, ed: *Advances in Econometrics and Modeling* (Kluwer Academic Publishers, Inc., Norwell, Mass.).

Pierse, R. G., and A. J. Snell, 1992, Temporal aggregation and the power of tests for a unit root, Unpublished manuscript, University of Cambridge.

Richardson, Matthew, and Tom Smith, 1991, Tests of financial models in the presence of overlapping observations, *Review of Financial Studies* 4, 227–254.

——— and James H. Stock, 1989, Drawing inferences from statistics based on multi-year asset returns, *Journal of Financial Economics* 25, 323–348.

Schwert, G. William, 1989, Why does stock market volatility change over time?, *Journal of Finance* 44, 1115–1153.

———, 1990, Stock returns and real activity: A century of evidence, *Journal of Finance* 45, 1237–1257.

Stambaugh, Robert F., 1986, Bias in regression with lagged stochastic regressors, CRSP Working Paper No. 156, University of Chicago.