William N. Goetzmann
Columbia University

Patterns in Three Centuries of Stock Market Prices*

This article applies autoregression and rescaled range statistics to very long stock market series to test the hypothesis that long-term temporal dependencies are present in financial data. For the annual capital appreciation returns to the London Stock Exchange, evidence of persistence in raw returns greater than 5 years and of mean reversion in deviations from rolling 20-year averages is found. Similar patterns are observed for the New York Stock Exchange; however, they are not significant at traditional confidence levels.

I. Introduction

Tests about the temporal behavior of long-horizon stock returns by Fama and French (1988) have suggested the possibility of mean reversion in stock prices and breathed new life into the Dow theory (see Rhea 1932), which claims that the stock market follows an alternating pattern of bull and bear markets. The importance of the Fama and French research lies not in the identification of a particular stochastic model of stock return behavior but in the implication explored by, among others, Poterba and Summers (1988) and Shiller (1989) that financial markets may be subject to temporary "fads" or at least periodically time-varying expected returns. The Fama and French research has motivated several methodological studies of mean reversion tests. Bootstrapping tests by Kim, Nelson, and Startz (1988), Goetzmann (1990), McQueen (1992), and Richardson (in press) have explicitly modeled serial independence of monthly and annual stock returns and have tended to reject mean reversion of long-term stock returns in favor of the more parsimonious random-walk model of multiple year stock returns.

* I wish to thank Roger Ibbotson, Stephen Ross, Jonathan Ingersoll, Jr., Robert Shiller, the editor, John Huizinga, and Kenneth R. French, the referee, for their comments. Thanks are also due to Christopher Musto for his research assistance and to Ibbotson Associates of Chicago for making their data available. I am solely responsible for all errors.

(*Journal of Business,* 1993, vol. 66, no. 2)

0021-9398/93/6602-0004$01.50

Poterba and Summers (1988) observe that the failure to distinguish between low-frequency mean reversion and complete unpredictability of returns lies in the power of the tests used to examine them. For instance, a test of mean reversion in 5-year stock market returns based on the data available from the Center for Research in Security Prices (CRSP) monthly files has only 11 independent observations—hardly enough to draw convincing conclusions about repeated temporal patterns.

This lack of data has led to ingenious attempts to extract more information from the existing series. Fama and French (1988) use overlapping observations rather than temporally independent returns and correct for the lack of independence in the errors by the method proposed by Hansen and Hodrick (1980). They find that 4- and 5-year returns to the equal-weighted New York Stock Exchange (NYSE) over the 1926–87 period are negatively autocorrelated. Richardson and Stock (1989) derive the distribution of the Fama and French regression coefficients and demonstrate that overlapping observations may contain more information than independent observations; however, they cannot reject the random walk model.

Another solution for increasing the power of the test is to collect more data. In tests of long-run price dependency, Poterba and Summers (1988) and Lo (1991) use the annual Cowles (1938) U.S. stock index extending back to 1872. Lo uses the rescaled range (R/S) statistic to test for aperiodic reversals and finds no evidence of them. Poterba and Summers use a variance ratio test and find marginal evidence of mean reversion, although the results are not strong enough to reject the random walk model at traditional confidence levels.

In this article, I extend market history even further back in time. Joint-stock shares have traded in London for 300 years and in New York for 200 years. If cycles of periodicity greater than a year are consistently present in British or American stock prices, one would expect to find them in the longest indices of all. These long series offer an opportunity to identify patterns that shorter time series cannot. For instance, tests of 5-year serial dependence in London and New York stock price indices using the available published data may employ 57 and 39 observations, respectively. In autocorrelation tests of mean reversion and persistence of multiple horizon capital appreciation returns, I apply the bootstrapping methodology to two stock price indices that extend back to the eighteenth century. I perform separate as well as joint significance tests with respect to models that hypothesize long-term periodic behavior. In addition, I measure the R/S statistic proposed by Mandelbrot (1972) and modified by Lo (1991) as a test of long-term dependency in prices. Contrary to the results from tests on the last 120 years of U.S. stock market data, the longer-term perspective suggests that the random walk model does not correctly describe the behavior of U.S. stock prices. I find evidence of long-term structural changes in stock price appreciation. Once these structural changes are incorporated in the tests, I find some evidence of a persistent mean-reverting component in stock market prices of the sort discussed by Fama and French (1988) and Poterba and Summers (1988). Although autocorrelation tests on the long-term NYSE capital appreciation index yield test statistics consistent with mean reversion, the null hypothesis of temporal independence cannot be rejected at traditional confidence levels. The R/S tests, however, provide some evidence that the detrended London Stock Exchange (LSE) and NYSE prices may exhibit long-term memory.

This article is organized as follows: Section II describes the sources of the data and discusses the possible errors and biases in each series. Section III reports the methodology and results of autoregression tests. Section IV reports the methodology and results of the rescaled range tests. Section V concludes.

II. Data Sources

Shares of the Bank of England were traded on the Royal Exchange in the seventeenth century, and British publications such as John Castaing's *The Course of the Exchange* regularly reported share prices of at least six joint stock companies, beginning in the eighteenth century.[1] According to Mirowski (1981), who compiled an equal-weighted average of British share prices through the eighteenth century, the London market for shares was active and fully functional by 1700, although the frequency of trades and price quotations fluctuated considerably. In this article, I analyze an annual share price index for the LSE compiled from seven different sources, beginning with the Mirowski (1981) index. These sources are reported and described in table 1. Gayer, Rostow, and Schwartz (1953) provide a broad-based index of shares through the first half of the nineteenth century and report the Haekel index that extends the LSE index to 1866. Several economists have constructed indices for periods of the late nineteenth and early twentieth centuries—I use Bowley, Schwartz, and Smith (1931) and Smith and Horne (1934). Two financial periodicals provide index measures of equity price appreciation through the early and middle twentieth century: the *Bankers Magazine* (1915–27) and the *Economist* (to 1970), which prints the *Financial Times Index*. Data on the capital appreciation of the LSE since 1970 are collected by the *Financial*

1. See Neal (1990) for a discussion of *The Course of the Exchange*. It was published semiannually, with daily price quotes for major stocks, over the period from 1698 to 1810. These are available in electronic form from Inter-University Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, MI 48106.

TABLE 1 Sources of Data for the Construction of the London Share Price Index

Dates	Source	Firms	Types of Firms	Mean	SD	Auto-correlation	Method and Possible Biases
1695–1809	Mirowski 1981	six or less	Banks, insurance, and trading	.015	.139	−.21	Equal weighted. Selected regularly quoted firms. Infrequent trading and survivorship bias are a problem.
1810–50	Gayer, Rostow, and Schwartz 1953	68 entered or exited index	Broad-based. Including banks, insurance, transportation, and mining and utilities	.008	.092	.29	Value weighted. Selected regularly quoted, representative firms. Some interpolation or smoothing as a result of infrequent trading. Survivorship bias minimal.
1851–66	Haekel Index, in Gayer, Rostow, and Schwartz 1953	unknown	Unknown	.017	.094	.00	Unknown
1867–1914	Smith and Horne 1934	from 25 to 77 companies	Broad-based. Including manufacturing, construction, retail, transportation, and communication. No banks or insurance.	.014	.069	.30	Equal weighted. Selected regularly quoted, representative firms. Split correction unclear. Survivorship bias minimal.
1915–27	*Bankers Magazine* 1915–27	more than 200 per year	Virtually all "variable dividend" securities quoted on the exchange	.011	.089	.00	Equal weighted. Survivorship bias minimal.
1928–29	Bowley, Schwartz, and Smith 1931	92 industrials	Broad-based. Including manufacturing, construction, retail, transportation, and banks	−.034	.135	.00	Value weighted. Selected regularly quoted "important" firms. Survivorship bias minimal.
1930–70	Financial Times Index, reported in *Economist*, 1930–70	30 industrials	"Blue-Chip" index, representing several industries	.065	.189	.00	Share-price weighted. Selected regularly quoted, "important" firms. Survivorship bias minimal.
1971–89	Financial Times Actuaries Index for 1971–89 (used with permission of Ibbotson Associates, Chicago)	500 industrial companies	Broad-based, representing all industries	.203	.355	−.37	Value weighted. Survivorship bias minimal.

Times in their *Financial Times Actuaries Index* and by Morgan Stanley Capital International. Both of these are available from Ibbotson Associates, Inc. Because the LSE index is spliced from many sources, it does not necessarily reflect the continuous performance of an investable portfolio through time, although it is probably a fair approximation. The existence of overlapping observations at each splice insures that there is no abrupt change in the composition of the index that could be misinterpreted as an actual return.

Active trading in shares in the United States dates from the end of the eighteenth century (see table 2). The New York Stock Exchange was founded in 1792, and the Foundation for the Study of Cycles (see Ibbotson and Brinson 1987) has compiled an annual stock price series from 1790 to the present that combines a number of other studies.[2] Unfortunately, many of the component series suffer from biases due to smoothing and survivorship. For the period from 1815 to 1859, a broad-based index of the New York Stock Exchange is available from Goetzmann and Ibbotson (1992). It is an equal-weighted annual index of all listed equity shares on the NYSE and is based on the price quotes in the *New York Shipping and Commercial* (see *New York Shipping List* [1815–1926]), which provided the official record of NYSE price quotes and representative transactions prices for several decades of the early nineteenth century. It deals with the problem of infrequent trading through the use of the weighted repeat sales method proposed by Case and Shiller (1987) and studied by Goetzmann (1992). For the period from 1860 to 1871, I must again rely on the Foundation for the Study of Cycles index, which is probably composed of the Cole-Frickey (1928) index of railroad shares over this period and is thus not broad based. The index created by Cowles (1938) and adjusted for data errors by Wilson and Jones (1987) begins in 1872 and is constructed using the average of the high and low prices of individual stocks in each month. It is a capital-weighted index that is broad based, but it may be subject to survivorship bias, and, as Working (1960) points out, the averaging procedure introduces monthly smoothing. After 1926, I use the capital appreciation return to the Standard and Poor's index, reported by Ibbotson Associates (1991). For a complete discussion of

2. See Ibbotson and Brinson (1987). They explain that the index is composed of "an internal index . . . the Cleveland Trust Company Index . . . the Clement-Burgess Index and the Cowles Index" (p. 73). The Cleveland Trust Company Index includes indices compiled from other sources—probably Cole and Frickey (1928), while the Clement Burgess Index is extremely narrow. As Cowles (1938) notes, it is "composed of from four to nine stocks, chiefly leading railroads" (p. 439). This narrow base, which covers periods in the nineteenth century when the NYSE listed over one hundred frequently traded stocks, suggests that the index may have been created by using only companies with data extending over the entire period of study, that is, 1854–83. Given this survivorship bias, the Foundation for the Study of Cycles NYSE index from 1790 to the late nineteenth century may be positively biased in the early years.

TABLE 2 Sources of Data for the Construction of the NYSE Share Price Index

Dates	Source	Firms	Types of Firms	Mean	SD	Auto-correlation	Method and Possible Biases
1790–1815	Foundation for the Study of Cycles, reprinted in Ibbotson and Brinson 1987	Probably less than 20	Banks and insurance	.076	.256	−.21	Not reported. Possibly spliced from other sources discussed in Schwert (1990).
1816–59	Goetzmann and Ibbotson 1992	20–260	Broad-based. Includes all of the firms listed on the NYSE over the period. For early years, these are principally banks and insurance companies. Later, they include transportation, mining and utilities, and industrials.	.034	.127	.29	Equal-weighted estimate using repeat-sale index construction methodology. May overstate variance and induce negative autocorrelation at 1-year intervals but not longer. May induce some smoothing as well. Survivorship bias minimal.
1860–71	Foundation for the Study of Cycles, reprinted in Ibbotson and Brinson 1987	Less than 20	Very narrow. Probably only frequently traded railroads.	.118	.180	.00	Probably based on the Cole and Frickey Index of Railroad Stocks, discussed in Schwert (1990). Extreme survivorship bias.
1872–1925	Cowles 1938; Wilson and Jones 1987	12–351	Broad-based. Including manufacturing, construction, retail, transportation, and communication.	.031	.161	.30	Value-weighted averages of monthly high and low prices inducing some time averaging. Survivorship bias minimal.
1926–89	S&P Index reported by Ibbotson Associates 1991	90–500	Broad-based. Including manufacturing, construction, retail, transportation, and communication.	.071	.200	.00	Value-weighted average of capital appreciation of industrial shares. Survivorship bias and time averaging are minimal.

the limitations and biases in the pre-CRSP U.S. stock return series, the reader is referred to Schwert (1990).

As noted, the problems associated with the spliced long-term appreciation series tend to bias estimates of both the long-term mean and the standard deviation and to a lesser extent the annual autocorrelation. The long-term mean may be upwardly biased due to the selection of frequently traded or surviving securities used to create the indices over the early periods. The standard deviation may fluctuate since the number of stocks in each series also fluctuates—because of the effect of diversification, one would expect the standard deviation of the indices to decline as the number of stocks increases. The direction of the autocorrelation bias at the annual horizon is not clear. While smoothing may be caused by averaging of high and low prices and by infrequent price observations, some annual negative correlation in the NYSE series may be induced by the use of the repeat-sales method. For horizons greater than 1 year, both effects decline in importance; however, the question of dividend yield becomes significant. If total returns are independent, but dividend policy changes slowly through time, then one would observe long-term price dependency. Consequently, in the following tests, I allow for fixed as well as slowly changing mean values. Unfortunately, there is no ready evidence for dividend yields from the early periods covered by the data.[3]

Perhaps more significant than the biases introduced by survivorship, recording methods, data splicing, and dividend policy changes is the fact that the economies of both countries changed profoundly over the course of the last 3 centuries. The LSE series, for instance, documents share prices through the entire industrial revolution, the nation's colonial expansion, and centuries of development in the capital markets. Similarly, the NYSE documents the U.S. equity market over the period of westward expansion, the development of the U.S. rail transportation system, and the evolution of the economy from agrarian to industrial. Such broad historical changes are reflected in the composition of both indices as different types of corporations financed growth through the equity markets. Not only would some of these firms have different expected returns, but they would also reflect different kinds of risks. These broad, evolutionary issues present problems in regression tests of mean reversion since the tests assume stationarity of the parameters of the model to be estimated. Thus, by gathering more data I have solved some problems, while introducing others.

Table 3 reports summary statistics for the NYSE and the LSE over

3. Indirect evidence for the NYSE stocks is implicit in the manner in which stock prices were quoted on the exchange. Prices were quoted with respect to a par value of 100, with prices rarely deviating above 200, and splits were practically nonexistent in the early nineteenth century. This suggests that investors expected earnings to be paid out rather than retained.

TABLE 3 Summary Statistics for LSE and NYSE Capital Appreciation Indices

	Arithmetic Mean	Geometric Mean	Standard Deviation	Skewness Log(1 + r)	Kurtosis Log(1 + r)	Auto-correlation
Over the entire length of each series:						
LSE	.031	.021	.157	.013	13.37	−.049
NYSE	.056	.039	.184	−.667	4.60	.017
Over the eighteenth century, through 1800:						
LSE	.014	.005	.144	−1.098	22.166	−.217
NYSE	.076	.066	.150	−.363	1.233	.184
Over the nineteenth century, 1801–1900:						
LSE	.015	.012	.085	−.256	3.263	.266
NYSE	.047	.032	.177	−.343	5.596	.069
Over the twentieth century, 1901–89:						
LSE	.069	.049	.218	.255	6.266	−.065
NYSE	.063	.043	.196	−.886	3.829	−.025

NOTE.—NYSE series for the eighteenth century begins in 1790. Skewness and kurtosis are estimated from the log of one plus the return, which is approximately normal. A Kolmogorov-Smirnov test of normality for each series fails to reject at the 95% confidence level. The probability of rejecting the null that LSE and the NYSE series are distributed log normally is 99.8% and 18.7%, respectively.

the entire period for which I have data and also breaks the results down by centuries. These summary figures are interesting in their own right. The long-term annual geometric capital appreciation return of the LSE, based on 290 years of data, is 2.1%. The long-term annual geometric capital appreciation return of the NYSE, based on 197 years of data, is 3.9%. The annual standard deviation of returns to the LSE and NYSE is 15.7% and 18.4%, respectively. The mean return in each country differs significantly over the nineteenth century but not so over the twentieth century. This may reflect the economic return of the underlying assets themselves, or it may reflect international differences in dividend yields. When logged, neither distribution is dramatically skewed, while both are leptokurtotic when compared to normal distribution. Despite the kurtosis, however, a Kolmogorov-Smirnov test rejects normality for the LSE but not the NYSE.[4]

Figure 1 plots the LSE and NYSE capital appreciation indices. It is clear that the variance of the LSE is not stable over time; note the high variance periods in the 1720s and in the 1950s through the 1980s, with a long stretch of relative calm in between. The NYSE variance is more stable, although the Great Depression stands out as a period of relatively high volatility, along with occasional dramatic outliers in the first 50 years of the series. Note also that the appreciation rate of the LSE increases dramatically over time. It exhibits practically no increase through the eighteenth century and appears to increase at a lower rate than the NYSE through the nineteenth century and the first half of the twentieth century. After 1950, it appears to increase at a greater rate than the NYSE.

III. Methodology

A. *Autocorrelation Tests*

Following Fama and French (1988), I use an autoregression of multiple year capital appreciation returns to test for long-term serial dependency in stock market prices. Without an ex ante hypothesis regarding the number of years to include in the compound returns, I test the serial dependency of 1–10-year horizon returns. That is,

$$r(t, t + T) = \alpha(T) + \beta(T)r(t - T, t) + e(t, t + T), \tag{1}$$

for $t = 0, T, 2T, \ldots, nT$, where $T = 1, \ldots, 10$.

Unlike Fama and French (1988), I use only nonoverlapping returns, so that I do not need to correct for serial dependency in the residuals;[5]

4. The probability associated with rejection of the null hypothesis that the series is drawn from a normal distribution is .997 for the LSE and .1872 for the NYSE.

5. Fama and French (1988) employ the Hansen and Hodrick (1980) correction for overlapping returns. Richardson and Smith (1991) have analyzed the behavior of statis-

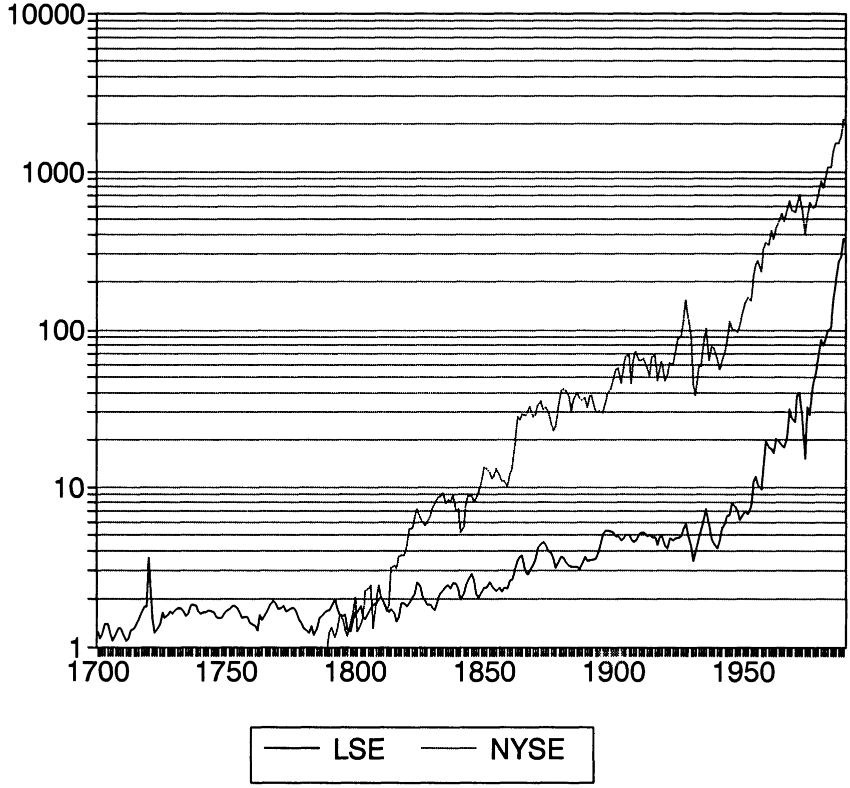

FIG. 1.—London and New York Stock Exchanges Capital Appreciation Indices, 1790–1989. The dotted line (begins ca. 1800) indicates the annual capital appreciation of the NYSE; the solid line (begins ca. 1700) indicates the annual capital appreciation of the LSE; the *X*-axis indicates the time period from 1790 to 1989; the *Y*-axis indicates the growth of an invested dollar or pound over the period. Sources for both indices are described in the text.

however, I correct for the bias in the autoregression coefficient noted by Kendall (1954) by bootstrapping the autoregression coefficient, under the null hypothesis that successive annual capital appreciation returns to the stock market are independent and identically distributed. The bootstrap is performed by drawing r^*, a bootstrapped pseudohistory of market returns with replacement from the empirical distribution of $r(t)$. The multiple year returns are formed by compounding $r^*(t)$, and the regression test is performed 1,000 times, providing a distribution of

tics in the presence of overlapping observations and devised appropriate corrections for hypothesis tests involving regression coefficients. Because I am interested in the explanatory power, as measured by R^2 as well as the significance of the regression coefficients, I have chosen to use nonoverlapping observations.

regression coefficients, β^*, and R^{2*}:

$$r^*(t, t+T) = \alpha^*(T) + \beta^*(T)r^*(t-T, t) + e^*(t, t+T). \qquad (2)$$

The bootstrap not only provides a correction for the autocorrelation bias, but it generates distributions of regression statistics that conform to the null hypothesis that returns are independently and identically distributed (i.i.d.). By comparing the values of β and R^2 to their bootstrapped distributions, I may determine how unusual they are, given a null hypothesis that successive annual returns are independent and identically distributed. In addition, the standard deviation of the bootstrapped distribution of the regression coefficients is a consistent estimate of the coefficient standard error (see Efron 1979) and is used to construct a *t*-statistic.

Since the dividend yield may have changed over the course of several centuries, I also perform a test that allows for long-term variation in the mean return. Instead of the total capital appreciation, I examine the deviation of the annual capital appreciation return from its 20-year moving average:

$$r^d(t) = r(t) - \frac{1}{20}\sum_{i=t-1}^{t-20} r(i). \qquad (3)$$

Because the dividend yields are unknown, I cannot distinguish between long-term changes in expected returns and changing dividend policies. However, for horizons less than 20 years, I can test whether deviations from the long-term mean are reverting or persistent.[6]

As noted earlier, there is no reason to expect that appreciation returns to 3 centuries of stock prices will be homoscedastic. In fact, since the series are spliced from components with varying numbers of securities, I expect the variance of different sections of each index to differ. Consequently, I perform a stratified-variance bootstrap, using the method proposed by Kim, Nelson, and Startz (1988). For both the raw and the demeaned series, I divide the sample into five groups, according to variance, where variance is defined by squared returns. I sample with replacement from each stratum in order to match the approximate temporal pattern of heteroscedasticity present in the sample.[7] As with the bootstrap draws under the assumption that returns

6. Given the fact that the mean is allowed to vary with time, the term "mean reversion" actually only applies to the deviations from the moving average. Indeed, if the mean were allowed to vary each period, a test of "mean reversion" would be absurd. Thus, the test of deviations from the long-term average should be interpreted as conditional upon the specification in eq. (3).

7. To the extent that the changing means are components of the squared returns, this will tend to make the bootstrap sample resemble the true sample in general. This weakens the power of the test, and thus I will report bootstrap quantiles generated by the i.i.d. and stratified variance procedures.

are i.i.d., I perform the autocorrelation tests, save the coefficients and R^2s, and then report the empirical quantiles exceeded by the statistics from tests performed on the actual series.

B. R/S Tests

Autocorrelation tests detect long-term dependency in stock market prices if the dependent behavior is periodic and if the periodicity is consistent over time. Fundamental historical changes may have altered the period of market cycles, however. Mandelbrot (1972) proposes a statistic to measure the degree of long-term dependency, in particular, "nonperiodic cycles." The rescaled range, or R/S statistic, is formed by measuring the range between the maximum and minimum distances that the cumulative sum of a stochastic random variable has strayed from its mean and then dividing this by its standard deviation. An unusually small R/S measure would be consistent with mean reversion, for instance, while an unusually large one would be consistent with return persistence.

Mandelbrot (1972) has shown that the R/S statistic is a more general test of long-term dependency in time series than either autocorrelation tests or examination of spectral densities. He points out that, in particular, it is robust to changes in periodicity. Lo (1991) points out that one limitation of the R/S statistic is that it cannot distinguish between short- and long-term dependency, nor is it robust to heteroscedasticity.

Lo (1991) modifies the R/S statistic so that it is more robust to violations in the assumption that returns are i.i.d. The modification consists of replacing the standard deviation with an estimate that explicitly models short-term temporal dependency using the autocovariances up to a finite number of lags, weighted by factors proposed by Newey and West (1987):

$$R/S_{\mathrm{Lo}} = \frac{1}{\sqrt{T\hat{\sigma}^*}}\left[\max_{1\leq\tau\leq T}\sum_{t=1}^{\tau}(r_t - \bar{r}) - \min_{1\leq\tau\leq T}\sum_{t=1}^{\tau}(r_t - \bar{r})\right], \tag{4}$$

where

$$\hat{\sigma}^{2*} = \hat{\sigma}^2 + 2\sum_{t=1}^{q}\omega_t(q)\,\hat{\gamma}_t,$$

$$\omega_t(q) \equiv 1 - \frac{t}{q+1},$$

and

$$\hat{\gamma}_t \equiv \text{the autocovariance operator.}$$

Lo (1991) points out that the uncorrected R/S statistic is sensitive to heteroscedasticity and cannot distinguish the compounded effects of short-horizon patterns from long-term patterns. He derives the distribution of the modified (R/S) statistic, allowing it to be used in a hypothesis test about long-term dependency in stock market returns.[8] While the R/S statistic identifies nonperiodic cycles, it is not free of the choice of return horizon. As with the autocorrelation test, I report the bootstrapped quantile exceeded by the R/S statistic for the raw stock series and the demeaned stock series, under the i.i.d. and stratified variance procedures. In addition, I report the exceeded quantile of the analytically derived distribution of the modified R/S reported in Lo (1991).

C. *Joint Hypothesis Tests*

One problem with examining either the autocorrelation coefficient or the R/S statistic for a number of different horizons is that a hypothesis test about the significance of a subset of the coefficients or R/S statistics is misleading. Thus, as in Goetzmann (1989), Richardson and Smith (1991), and Richardson (in press), I perform a joint significance test across all 10 autoregression coefficients. To test that all 10 coefficients are zero, I use the Wald test:

$$W(\overline{\beta}) \equiv T(\overline{\beta}\Omega^{-1}\overline{\beta}) \sim \chi^2_k, \tag{5}$$

where $\overline{\beta}$ denotes the bias-corrected coefficient vector, and the covariance matrix that describes the cross-horizon dependencies is estimated with the bootstrap

$$\hat{\Omega}_{i,j} = \hat{\sigma}^2_{(\beta_i^*,\beta_j^*)}. \tag{6}$$

The χ^2 distribution is known to be sensitive to deviations from normality in the underlying distribution. Thus the parametric Wald test may be misspecified. Fortunately, the Wald statistic of equation (5) suggests a nonparametric test as well. A rejection region for the W-statistic based on the distribution of the bootstrapped Wald statistic, W^*, may be identified. In other words, I calculate the Wald statistic for each bootstrapped coefficient vector and use the resulting distribution for hypothesis testing. Thus, while a comparison of the Wald statistic to the χ^2 distribution may cause the null to be rejected as a result of deviations of the coefficient vector from multivariate normality, a comparison to the bootstrapped distribution of the Wald statistic will not since it is based on draws from the empirical coefficient distribution.

As with the autocorrelation test, it is known that the R/S statistic for each horizon is not independent. Thus, it is necessary to perform a joint test of dependency across all 10 lags as before. Lo (1991) demonstrates that the distribution of the R/S statistics is defined by the

8. Green and Fielitz (1977) applied the R/S statistic to examine U.S. stock market behavior but did not formulate an explicit test of long-term market memory.

range of a Brownian bridge process—a distribution that is slightly right-skew and leptokurtotic in comparison to the normal. This biases the Wald test toward rejection; however, as I noted above, a comparison of the Wald statistic to its bootstrapped distribution, rather than the χ^2 distribution, is robust to departures from normality. It may thus be used as a measure of whether R/S statistics for each horizon are jointly unusual.

IV. Results

Table 4 and table 5 report the autocorrelation and R/S tests for the raw LSE series and the demeaned LSE series, respectively. The bootstrap t-statistic and the bootstrap percentiles derived from the i.i.d. and stratified variance methods are reported for each autoregression coefficient at return horizons from 1 to 10 years.[9] I report bootstrap percentiles for the R/S statistic, as well as Lo's (1991) analytically derived percentiles. In addition, I report the autoregression R^2 with bootstrapped percentiles.

Table 4 suggests that returns with horizons greater than 5 years are strongly persistent. The bootstrapped percentiles indicate that the 6-, 8-, 9-, and 10-year coefficients are strongly positive and that the explanatory power of the regression, as measured by R^2, is around 10%–20%. This result may be partially due to the variance structure of the time series, however. Only the 8-year coefficient exceeds the ninety-fifth percentile of the stratified variance coefficient distribution. The R/S statistics are unusual at the first three horizons but not at longer intervals. Since they exceed one for horizons up to 3 years, they suggest persistence rather than reversion.[10]

Table 5 reports the results from the demeaned LSE series. It shows that much of the persistence in raw LSE returns may in fact be due to long-term changes in the mean. Once the 20-year moving average is subtracted, all of the coefficients become negative, and for the 4-, 5-, 6-, and 7-year horizons they are significantly so. In fact, the coefficients display the U-shaped pattern that Fama and French (1988) predict for returns generated by a process having both permanent and temporary components. The evidence from the R/S statistics is less clear. Few are unusual when compared to Lo's analytically derived distribution or to the bootstrapped distributions.

Table 6 and table 7 report the autocorrelation and R/S tests for the

9. The bootstrap t-statistic is formed by dividing the bias-adjusted coefficient vector by the standard deviation of the bootstrapped coefficient samples, in the manner proposed by Efron (1979).

10. The fact that the R/S statistic does not exceed one for longer horizons does not imply a contradiction between the autocorrelation test and the R/S test since Lo's (1991) adjustment of the R/S statistic includes multiple lags.

TABLE 4 **Autoregression and Rescaled Range Statistics for Multiple-Year Capital Appreciation Returns to the London Stock Exchange Index, 1700–1989**

Horizon (Years)	Number of Observations	Coefficient*	R/S Lo Statistic†	R^2†
1	289	−.049	1.953	.002
		t = 3.26, p = .20, q = .45	lo = .97, p = .99, q = .49	p = .67, q = .50
2	144	.093	1.713	.008
		t = 3.25, p = .91, q = .36	lo = .90, p = .99, q = .33	p = .79, q = .35
3	95	−.125	1.764	.011
		t = 1.15, p = .11, q = .19	lo = .95, p = .99, q = .59	p = .75, q = .71
4	71	−.01	1.456	.000
		t = 1.45, p = .53, q = .52	lo = .70, p = .93, q = .32	p = .16, q = .05
5	57	.025	1.318	.000
		t = 1.72, p = .63, q = .78	lo = .60, p = .83, q = .29	p = .22, q = .15
6	47	.543	1.193	.142
		t = 5.58, p = 1.00, q = .79	lo = .40, p = .63, q = .13	p = .99, q = .56
7	40	.052	1.317	.002
		t = 1.49, p = .71, q = .25	lo = .60, p = .90, q = .59	p = .29, q = .20
8	35	1.00	1.106	.227
		t = 6.54, p = 1.00, q = 1.00	lo = .30, p = .45, q = .19	p = 1.00, q = .77
9	31	.681	1.023	.097
		t = 3.22, p = 1.00, q = .87	lo = .20, p = .26, q = .06	p = .95, q = .59
10	28	.929	1.053	.194
		t = 3.89, p = 1.00, q = .93	lo = .20, p = .31, q = .27	p = .99, q = .39

NOTE.—Autoregressions are estimated from the model given in eq. (1). Lo's rescaled range statistics are heteroscedasticity-consistent after Lo (1991) and are estimated from the model given in eq. (4). Quantiles for the R/S statistic from Lo (1991) are indicated by "lo." t-statistics are bias corrected using the median values of the bootstrapped distributions and use the standard deviation of the bootstrap coefficient distribution. Probability values p indicate the bootstrap distribution percentile exceeded by the statistic in 1,000 iterations. Probability values q indicate the stratified-variance bootstrap distribution percentile exceeded by the statistic in 1,000 iterations.

* Bootstrap t-statistic and percentiles for i.i.d. and stratified variance draws are also reported.

† Bootstrap percentiles for i.i.d. and stratified variance are also reported.

TABLE 5 **Autoregression and Rescaled Range Statistics for Multiple-Year Capital Appreciation Returns to the London Stock Exchange Index, 1700–1989: Corrected for Long-Term Changing Means**

Horizon (Years)	Number of Observations	Coefficient*	R/S Lo Statistic†	R^2†
1	269	−.136	1.144	.02
		$t = 1.06, p = .01, q = .09$	lo = .30, $p = .54, q = .24$	$p = .98, q = .90$
2	134	−.115	.926	.01
		$t = .39, p = .08, q = .00$	lo = .05, $p = .15, q = .02$	$p = .82, q = .95$
3	89	−.232	.943	.06
		$t = 1.46, p = .00, q = .08$	lo = .10, $p = .16, q = .09$	$p = .97, q = .91$
4	66	−.372	.925	.13
		$t = 3.64, p = .00, q = .02$	lo = .05, $p = .12, q = .01$	$p = 1.00, q = .98$
5	53	−.375	1.225	.13
		$t = 3.25, p = .00, q = .04$	lo = .50, $p = .65, q = .88$	$p = .99, q = .96$
6	44	−.399	1.060	.15
		$t = .60, p = .00, q = .02$	lo = .20, $p = .29, q = .47$	$p = .99, q = .98$
7	37	−.358	1.217	.13
		$t = .60, p = .01, q = .00$	lo = .50, $p = .60, q = .79$	$p = .99, q = 1.00$
8	32	−.260	1.368	.07
		$t = .78, p = .06, q = .07$	lo = .70, $p = .86, q = .98$	$p = .89, q = .90$
9	29	−.183	.871	.04
		$t = 1.42, p = .15, q = .04$	lo = .05, $p = .02, q = .08$	$p = .71, q = .65$
10	26	−.194	.959	.04
		$t = 1.39, p = .17, q = .04$	lo = .10, $p = .09, q = .40$	$p = .67, q = .60$

NOTE.—Autoregressions are estimated from the model given in eq. (1). Rescaled range statistics are estimated from the model given in eq. (4). Quantiles for the statistic from Lo (1991) are indicated by "lo." t-statistics are heteroscedasticity consistent, after White (1980), and bias corrected using the median values of the bootstrapped distributions. Probability values p indicate the bootstrap distribution percentile exceeded by the statistic in 1,000 iterations. Probability values q indicate the stratified-variance bootstrap distribution percentile exceeded by the statistic in 1,000 iterations.

* Bootstrap t-statistic and percentiles for i.i.d. and stratified variance draws are also reported.

† Bootstrap percentiles for i.i.d. and stratified variance draws are also reported.

TABLE 6 **Autoregression and Rescaled Range Statistics for Multiple-Year Capital Appreciation Returns to the New York Stock Exchange Index, 1790–1989**

Horizon (Years)	Number of Observations	Coefficient*	R/S Lo Statistic†	R^2†
1	199	.017 t = 4.42, p = .63, q = .75	1.190 lo = .40, p = .59, q = .83	.000 p = .19, q = .39
2	99	−.279 t = .57, p = .00, q = .06	1.349 lo = .60, p = .82, q = .95	.077 p = .99, q = .95
3	65	−.262 t = 1.93, p = .02, q = .44	1.358 lo = .60, p = .83, q = .87	.068 p = .97, q = .59
4	49	−.083 t = 1.97, p = .32, q = .41	1.372 lo = .60, p = .88, q = .91	.007 p = .45, q = .56
5	39	−.183 t = 3.13, p = .15, q = .32	1.431 lo = .70, p = .92, q = .55	.031 p = .75, q = .69
6	32	−.141 t = .85, p = .28, q = .51	1.246 lo = .50, p = .69, q = .75	.019 p = .56, q = .49
7	27	−.202 t = .11, p = .18, q = .22	1.600 lo = .80, p = .97, q = .98	.037 p = .70, q = .77
8	24	−.292 t = 1.44, p = .08, q = .64	1.386 lo = .70, p = .85, q = .81	.072 p = .84, q = .40
9	21	−.006 t = .97, p = .59, q = .81	1.395 lo = .70, p = .81, q = .68	.000 p = .03, q = .03
10	19	−.327 t = 2.87, p = .09, q = .81	1.201 lo = .40, p = .08, q = .49	.082 p = .54, q = .27

NOTE.—Autoregressions are estimated from the model given in eq. (1). Rescaled range statistics are estimated from the model given in eq. (4). Quantiles from the distribution given in Lo (1991) are indicated as "lo." t-statistics are bias corrected using the median values of the bootstrapped distributions and use the standard deviation of the bootstrap coefficient distribution. Probability values p indicate the bootstrap distribution percentile exceeded by the statistic in 1,000 iterations. Probability values q indicate the stratified-variance bootstrap distribution percentile exceeded by the statistic in 1,000 iterations.

* Bootstrap t-statistic and percentiles for i.i.d. and stratified variance draws are also reported.

† Bootstrap percentiles for i.i.d. and stratified variance draws are also reported.

TABLE 7 **Autoregression and Rescaled Range Statistics for Multiple-Year Capital Appreciation Returns to the New York Stock Exchange Index, 1790–1989: Corrected for Long-Term Changing Means**

Horizon (Years)	Number of Observations	Coefficient*	R/S Lo Statistic†	R^2†
1	179	.063	.710	.004
		t = .64, p = .77, q = .62	lo = .00, p = .23, q = .04	p = .55, q = .63
2	89	−.173	.783	.030
		t = .09, p = .04, q = .14	lo = .00, p = .24, q = .04	p = .80, q = .86
3	59	−.230	.819	.054
		t = .39, p = .05, q = .29	lo = .03, p = .24, q = .06	p = .81, q = .74
4	44	−.067	.970	.005
		t = 3.43, p = .35, q = .39	lo = .10, p = .33, q = .04	p = .30, q = .53
5	35	−.167	1.175	.026
		t = 3.54, p = .20, q = .24	lo = .40, p = .60, q = .03	p = .57, q = .76
6	29	−.140	1.761	.019
		t = .279, p = .27, q = .11	lo = .95, p = .99, q = .83	p = .46, q = .88
7	24	−.253	1.667	.063
		t = .04, p = .13, q = .08	lo = .90, p = .97, q = .92	p = .71, q = .91
8	21	−.501	2.369	.236
		t = .81, p = .00, q = .06	lo = .99, p = .99, q = 1.00	p = .88, q = .95
9	19	−.239	1.515	.050
		t = 3.57, p = .15, q = .19	lo = .80, p = .89, q = .94	p = .61, q = .75
10	17	−.567	.976	.248
		t = 4.05, p = .00, q = .03	lo = .10, p = .29, q = .15	p = .87, q = .97

NOTE.—Autoregressions are estimated from the model given in eq. (1). Rescaled range statistics are estimated from the model given in eq. (4). t-statistics are bias corrected using the median values of the bootstrapped distributions and use the standard deviation of the bootstrap coefficient distribution. Probability values p indicate the bootstrap distribution percentile exceeded by the statistic in 1,000 iterations. Probability values q indicate the stratified-variance bootstrap distribution percentile exceeded by the statistic in 1,000 iterations.

* Bootstrap t-statistic and percentiles for i.i.d. and stratified variance draws are also reported.

† Bootstrap percentiles for i.i.d. and stratified variance draws are also reported.

raw NYSE and demeaned NYSE series, respectively. The coefficients for both series are negative at each horizon, although they do not follow the U shape hypothesized by Fama and French (1988). The *t*-statistics and the bootstrap probability levels suggest that some bias-adjusted coefficients may differ significantly from zero. Table 7 reports the results for the demeaned NYSE series. As with the demeaned LSE series, long-horizon returns appear significantly negatively autocorrelated. In addition, the R/S shows evidence of reversion at horizons less that 4 years.

While the statistics about the individual horizons are suggestive of mean-reverting behavior, the joint tests reported in table 8 indicate

TABLE 8 Joint Hypothesis Test for 1–10 Horizons Based on the Wald Statistic Described in Equation (5)

	Test Statistic	Quantile	90%	95%	99%
χ^2 distribution			15.99	18.31	23.21
LSE raw capital appreciation series autocorrelation coefficients:					
i.i.d. bootstrap	100.43	1.00	16.76	19.65	29.80
Stratified variance	38.41	1.00	16.20	19.12	26.13
LSE demeaned capital appreciation series autocorrelation coefficients:					
i.i.d. bootstrap	13.20	.79	16.32	19.45	28.07
Stratified variance	16.57	.90	16.02	18.57	24.07
NYSE raw capital appreciation series autocorrelation coefficients:					
i.i.d. bootstrap	14.37	.84	16.99	18.15	24.73
Stratified variance	8.73	.44	16.06	18.13	21.42
NYSE demeaned capital appreciation series autocorrelation coefficients:					
i.i.d. bootstrap	16.11	.90	15.48	18.26	24.09
Stratified variance	7.48	.31	15.97	18.21	23.54
LSE raw capital appreciation series R/S statistics:					
i.i.d. bootstrap	35.08	.98	19.02	23.83	41.18
Stratified variance	6.93	.32	16.90	21.32	28.10
LSE demeaned capital appreciation series R/S statistics:					
i.i.d. bootstrap	34.57	.98	16.22	22.66	57.56
Stratified variance	45.06	1.00	17.84	24.30	38.89
NYSE raw capital appreciation series R/S statistics:					
i.i.d. bootstrap	7.65	.68	14.76	21.65	61.59
Stratified variance	6.20	.34	16.00	19.18	38.47
NYSE demeaned capital appreciation series R/S statistics:					
i.i.d. bootstrap	32.84	.96	16.78	26.39	72.59
Stratified variance	34.73	.98	16.73	22.82	56.47

how unusual the pattern of 10 coefficients and R/S range statistics actually may be. The random walk is rejected under both the i.i.d. and the stratified variance procedures for the LSE at the 99% confidence level, using the autoregression coefficient test. The hypothesis that the NYSE autoregression coefficient vector is different from zero cannot be rejected, however. The rejection level of 84% under the i.i.d. sampling scheme is similar to rejection levels found by previous researchers using NYSE data over later periods. The joint tests on autoregression coefficients performed on the deviations from rolling 20-year means fail to reject the null at traditional confidence levels. Thus, while the results are suggestive of mean reversion in both markets, they are not conclusive when autoregression tests are used.

The joint tests on the more general R/S statistics yield slightly stronger results, however. While the joint tests performed on the R/S statistics derived from the raw LSE series and the raw NYSE series are inconclusive, the joint tests performed on the demeaned LSE and NYSE capital appreciation series are both significant at the 95% level. They indicate the likelihood that deviations from the lagged 20-year mean are not temporally independent. Table 5 and table 7 suggest the reasons for the joint departure from the null. In both tables, the R/S statistics are unusually low over 2–4-year horizons and unusually high over the 8-year horizon, regardless of whether the i.i.d. or the stratified variance bootstrap is used. Thus, the joint rejection does not result from a consistent deviation in one direction but from what appears to be reversion over the short horizons and persistence over the long horizons—even after the 20-year moving average has been removed.

V. Conclusion

The same tests used in previous research to demonstrate the lack of long-term memory in NYSE stock market prices during the various periods from 1872 to 1987 suggest that long-term memory may exist in LSE stock prices over the period of 1700–1989 and in deviations from 20-year means in both markets. This conclusion is based on autoregression tests as well as on R/S range tests and is robust to techniques designed to preserve the particular temporal pattern of stock market variance. These results may be interpreted as evidence of evolving dividend policies and/or changing expected financial returns that result from the changing composition of the index through the centuries of U.S. and U.K. capital market history. Substituting deviations from the lagged 20-year mean appears to eliminate the evolutionary effects from the LSE series, indicating that the long-term persistent patterns appear to mask a tendency for reversion toward the mean in the LSE and possibly in the NYSE. This behavior is consistent with models of stock returns proposed by Poterba and Summers (1988) and empirically examined by Fama and French (1988). It could be caused by rational time variation in expected returns, as postulated by Conrad and Kaul (1988) and explored by Jacquier and Nanda (1989) or by speculative bubbles of the sort discussed by DeBont and Thaler (1985) and Flood, Hodrick, and Kaplan (1987). Although consistent with all models that hypothesize long-term reversion in asset prices, the tests under discussion in this article, as currently formulated, cannot distinguish among them. Whether the serial dependence in long-term capital appreciation returns may be used to obtain arbitrage profits is another matter entirely. Rhea (1932) and his predecessor, Joseph Henry Dow, apparently thought so. A test of market efficiency based on the temporal patterns identified in this article would require a trading test and a measure of the total investor return rather than the capital appreciation component alone, and, in all likelihood, an investor horizon greater than a single lifespan.

References

Bankers Magazine (London). 1915–27. Available from 1844 to the present.

Bowley, A. L.; Schwartz, G. L.; and Smith, K. C. 1931. *A New Index of Securities.* Special Memorandum no. 33. London: London and Cambridge Economic Service.

Case, Karl, and Shiller, Robert. 1987. Prices of single family homes since 1970: New indexes for four cities. *New England Economic Review* (September/October), pp. 45–56.

Cole, Arthur H., and Frickey, E. 1928. The Course of Stock Prices, 1825–1866. *Review of Economics and Statistics* 10:117–39.

Conrad, Jennifer, and Kaul, G. 1988. Time variation in expected stock returns. *Journal of Business* 61:409–25.

The Course of the Exchange. 1698–1810. London: John Castaing and various publishers.

Cowles, Alfred, III. 1938. *Common Stock Indices, 1871–1937.* Cowles Commission for Research in Economics. Monograph no. 3. Bloomington, Ind.: Principia Press.

DeBondt, Werner, and Thaler, Richard. 1985. Does the stock market overreact? *Journal of Finance* 60:793–805.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7:1–26.

Fama, E. F., and French, K. R. 1988. Permanent and temporary components of stock prices. *Journal of Political Economy* 96:246–73.

Flood, K. R.; Hodrick, R.; and Kaplan, P. 1987. An evaluation of recent evidence of stock market bubbles. Working Paper no. 1971. Cambridge, Mass.: National Bureau of Economic Research.

Gayer, Arthur; Rostow, W. W.; and Schwartz, A. 1953. *The Growth and Fluctuation of the British Economy*. Vol. 1. Oxford: Clarendon Press.

Goetzmann, William N. 1990. Bootstrapping and simulation tests of long-term stock market efficiency. Ph.D. dissertation, Yale University.

Goetzmann, William N. 1992. The accuracy of real estate indices: Repeat sale estimators. *Journal of Real Estate Finance and Economics* 5:5–53.

Goetzmann, William N., and Ibbotson, Roger G. 1992. A broad-based index of New York Stock Exchange prices: 1815–1859. Working paper. New York: Columbia Business School.

Greene, Myron, and Fielitz, Bruce D. 1977. Long term dependence in common stock returns. *Journal of Financial Economics* 4, no. 3:339–49.

Hansen, Lars, and Hodrick, R. J. 1980. Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88:829–1054.

Ibbotson, Roger G., and Brinson, Gary. 1987. *Investment Markets*. Englewood Cliffs, N.J.: McGraw-Hill.

Ibbotson Associates. 1991. *Stocks, Bonds, Bills and Inflation: 1990 Yearbook*. Chicago: Ibbotson Associates, Inc.

Jacquier, Eric, and Nanda, V. 1989. Cyclicality of stock returns and the mean reversion puzzle. Working paper. Los Angeles: University of Southern California.

Kendall, M. 1954. A note on the bias in the estimation of autocorrelation. *Biometrika* 41:403–4.

Kim, Myung Jig; Nelson, C. R.; and Startz, R. 1988. Mean reversion in stock prices? Mimeographed. Seattle: University of Washington.

Lo, A. 1991. Long term memory in stock prices. *Econometrica* 59:1279–1314.

McQueen, Grant. 1992. Long-horizon mean-reverting stock prices revisited. *Journal of Financial and Quantitative Analysis* 27, no. 1 (March): 1–18.

Mandelbrot, B. 1972. Statistical methodology for non-periodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement* 1:259–90.

Mirowski, Phillip. 1981. The rise (and retreat) of a market: English joint stock shares in the eighteenth century. *Journal of Economic History* 41, no. 3:559–77.

Neal, Larry. 1990. *The Rise of Financial Capitalism*. Cambridge: Cambridge University Press.

Newey, W., and West, K. 1987. A simple positive definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703–5.

New York Shipping List, later *New York Shipping and Commercial*. 1815–1926. New York: Day & Turner. Volumes from 1816 in Beineke Rare Book Library, Yale University, New Haven, Conn.

Poterba, James M., and Summers, Lawrence H. 1988. Mean reversion in stock prices, evidence and implications. *Journal of Financial Economics* 22:27–59.

Rhea, Robert. 1932. *The Dow Theory*. New York: Barron's.

Richardson, Matthew. In press. Temporary components of stock prices: A skeptic's view. *Journal of Business Economics and Statistics*.

Richardson, Matthew, and Smith, Tom. 1991. Tests of financial models in the presence of overlapping observations. *Review of Financial Studies* 4, no. 2:227–54.

Richardson, Matthew, and Stock, James. 1989. Drawing inferences for statistics based on multiyear asset returns. Unpublished manuscript. Philadelphia: University of Pennsylvania, Wharton School.

Schwert, William. 1990. Indexes of U.S. stock prices from 1802 to 1897. *Journal of Business* 63, no. 3:399–426.

Shiller, Robert. 1989. *Market Volatility*. Cambridge, Mass.: MIT Press.

Shiller, Robert J., and Perron, P. 1985. Testing the random walk hypothesis: Power versus frequency of observation. *Economics Letters* 18:381–86.

Smith, K. C., and Horne, G. F. 1934. *An Index Number of Securities, 1867–1914*. Special Memorandum no. 37. London: London and Cambridge Economic Service.

White, Halbert. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817–38.

Wilson, Jack, and Jones, C. P. 1987. A comparison of annual common stock returns: 1871–1925 with 1926–85. *Journal of Business* 60:239–58.

Working, H. 1960. A note on the correlation on first differences of averages in random chains. *Econometrica* 28 (October): 916–18.