

# 6

## Long-Term Memory in Stock Market Prices

### 6.1 Introduction

THAT ECONOMIC TIME SERIES can exhibit long-range dependence has been a hypothesis of many early theories of the trade and business cycles. Such theories were often motivated by the distinct but nonperiodic cyclical patterns that typified plots of economic aggregates over time, cycles of many periods, some that seem nearly as long as the entire span of the sample. In the frequency domain such time series are said to have power at low frequencies. So common was this particular feature of the data that Granger (1966) considered it the “typical spectral shape of an economic variable.” It has also been called the “Joseph Effect” by Mandelbrot and Wallis (1968), a playful but not inappropriate biblical reference to the Old Testament prophet who foretold of the seven years of plenty followed by the seven years of famine that Egypt was to experience. Indeed, Nature’s predilection towards long-range dependence has been well-documented in hydrology, meteorology, and geophysics, and to the extent that the ultimate sources of uncertainty in economics are natural phenomena like rainfall or earthquakes, we might also expect to find long-term memory in economic time series.<sup>1</sup>

The presence of long-memory components in asset returns has important implications for many of the paradigms used in modern financial economics. For example, optimal consumption/savings and portfolio decisions may become extremely sensitive to the investment horizon if stock returns were long-range dependent. Problems also arise in the pricing of derivative securities (such as options and futures) with martingale methods, since the continuous time stochastic processes most commonly employed are incon-

<sup>1</sup>Haubrich (1993) and Haubrich and Lo (1989) provide a less fanciful theory of long-range dependence in economic aggregates.

sistent with long-term memory (see Maheswaran, 1990; Maheswaran and Sims, 1990; Sims, 1984, for example). Traditional tests of the capital asset pricing model and the arbitrage pricing theory are no longer valid since the usual forms of statistical inference do not apply to time series exhibiting such persistence. And the conclusions of more recent tests of "efficient" markets hypotheses or stock market rationality also hang precariously on the presence or absence of long-term memory.<sup>2</sup>

Among the first to have considered the possibility and implications of persistent statistical dependence in asset returns was Mandelbrot (1971). Since then, several empirical studies have lent further support to Mandelbrot's findings. For example, Greene and Fielitz (1977) claim to have found long-range dependence in the daily returns of many securities listed on the New York Stock Exchange. More recent investigations have uncovered anomalous behavior in long-horizon stock returns;<sup>3</sup> alternately attributed to speculative fads and to time-varying conditional expected returns, these long-run swings may be further evidence of the Joseph effect.

In this chapter we develop a test for such forms of long-range dependence using a simple generalization of a statistic first proposed by the English hydrologist Harold Edwin Hurst (1951). This statistic, called the "rescaled range" or "range over standard deviation" or " $R/S$ " statistic, has been refined by Mandelbrot (1972, 1975) and others in several important ways (see, for example, Mandelbrot and Taqqu, 1979, and Mandelbrot and Wallis, 1968, 1969a–c). However, such refinements were not designed to distinguish between short-range and long-range dependence (in a sense to be made precise below), a severe shortcoming in applications of  $R/S$  analysis to recent stock returns data since Lo and MacKinlay (1988b, 1990b) show that such data display substantial short-range dependence. Therefore, to be of current interest, any empirical investigation of long-term memory in stock returns must first account for the presence of higher frequency autocorrelation.

By modifying the rescaled range appropriately, we construct a test statistic that is robust to short-range dependence, and derive its limiting distribution under both short-range and long-range dependence. Contrary to the findings of Greene and Fielitz (1977) and others, when this statistic is applied to daily and monthly stock return indexes over several different sample periods and sub-periods, there is no evidence of long-range dependence once the effects of short-range dependence are accounted for. Monte Carlo experiments indicate that the modified  $R/S$  test has reasonable power against at least two particular models of long-range dependence, suggesting

<sup>2</sup>See LeRoy (1989) and Merton (1987) for excellent surveys of this recent literature.

<sup>3</sup>See, for example, Fama and French (1988), Jegadeesh (1989, 1990), and Poterba and Summers (1988).

that the time series behavior of stock returns may be adequately captured by more conventional models of short-range dependence.

The particular notions of short-term and long-term memory are defined in Section 6.2 and some illustrative examples are given. The test statistic is presented in Section 6.3 and its limiting distributions under the null and alternative hypotheses are derived via functional central limit theory. In Section 6.4 the empirical results are reported, and Monte Carlo simulations that illustrate the size and power of the test in finite samples are presented in Section 6.5. We conclude in Section 6.6.

## 6.2 Long-Range Versus Short-Range Dependence

To develop a method for detecting long-term memory, the distinction between long-range and short-range statistical dependence must be made precise. One of the most widely used concepts of short-range dependence is the notion of “strong-mixing” due to Rosenblatt (1956), a measure of the decline in statistical dependence between events separated by successively longer spans of time. Heuristically, a time series is strong-mixing if the maximal dependence between events at any two dates becomes trivially small as the time span between those two dates increases. By controlling the rate at which the dependence between past and future events declines, it is possible to extend the usual laws of large numbers and central limit theorems to dependent sequences of random variables. We adopt strong-mixing as an operational definition of short-range dependence in the null hypothesis of Section 6.2.1. In Section 6.2.2, we give examples of alternatives to short-range dependence such as the class of fractionally-differenced processes proposed by Granger and Joyeux (1980), Hosking (1981), and Mandelbrot and Van Ness (1968).

### 6.2.1 The Null Hypothesis

Let  $P_t$  denote the price of an asset at time  $t$  and define  $X_t \equiv \log P_t - \log P_{t-1}$  to be the continuously compounded single-period return of that asset from  $t-1$  to  $t$ . With little loss in generality, let all dividend payments be reinvested in the asset so that  $X_t$  is indeed the total return of the asset between  $t-1$  and  $t$ .<sup>4</sup> It is assumed throughout that

$$X_t = \mu + \epsilon_t, \quad (6.2.1)$$

where  $\mu$  is an arbitrary but fixed parameter and  $\epsilon_t$  is a zero mean random variable. Let this stochastic process  $\{X_t(\omega)\}$  be defined on the probability

<sup>4</sup>This is in fact how the stock returns data are constructed.

space  $(\Omega, \mathcal{F}, P)$  and define

$$\alpha(\mathcal{A}, \mathcal{B}) \equiv \sup_{\{A \in \mathcal{A}, B \in \mathcal{B}\}} |P(A \cap B) - P(A)P(B)|, \quad \mathcal{A} \subset \mathcal{F}, \mathcal{B} \subset \mathcal{F}. \quad (6.2.2)$$

The quantity  $\alpha(\mathcal{A}, \mathcal{B})$  is a measure of the dependence between the two  $\sigma$ -fields  $\mathcal{A}$  and  $\mathcal{B}$  in  $\mathcal{F}$ . Denote by  $\mathcal{B}_s^t$  the Borel  $\sigma$ -field generated by  $\{X_s(\omega), \dots, X_t(\omega)\}$ , i.e.,  $\mathcal{B}_s^t \equiv \sigma(X_s(\omega), \dots, X_t(\omega)) \subset \mathcal{F}$ . Define the coefficients  $\alpha_k$  as

$$\alpha_k \equiv \sup_j \alpha(\mathcal{B}_{-\infty}^j, \mathcal{B}_{j+k}^\infty). \quad (6.2.3)$$

Then  $\{X_t(\omega)\}$  is said to be *strong-mixing* if  $\lim_{k \rightarrow \infty} \alpha_k = 0$ .<sup>5</sup> Such mixing conditions have been used extensively in the recent literature to relax the assumptions that ensure the consistency and asymptotic normality of various econometric estimators (see, for example, Chan and Wei, 1988; Phillips, 1987; White, 1980; White and Domowitz, 1984). As Phillips (1987) observes, these conditions are satisfied by a great many stochastic processes, including all Gaussian finite-order stationary ARMA models. Moreover, the inclusion of a moment condition also allows for heterogeneously distributed sequences, an especially important extension in view of the apparent instabilities of financial time series.

In addition to strong mixing, several other conditions are required as part of the null hypothesis in order to develop a sampling theory for the test statistic proposed in Section 6.3. In particular, the null hypothesis is composed of the following four conditions on  $\epsilon_t$ :

(A1)  $E[\epsilon_t] = 0$  for all  $t$ ;

(A2)  $\sup_t E[|\epsilon_t|^\beta] < \infty$  for some  $\beta > 2$ ;

(A3)  $0 < \sigma^2 = \lim_{n \rightarrow \infty} E \left[ \frac{1}{n} \left( \sum_{j=1}^n \epsilon_j \right)^2 \right] < \infty$ ;

(A4)  $\{\epsilon_t\}$  is strong-mixing with mixing coefficients  $\alpha_k$  that satisfy

$$\sum_{j=1}^{\infty} \alpha_j^{1-\frac{2}{\beta}} < \infty.$$

Condition (A1) is standard. Conditions (A2) through (A4) are restrictions on the maximal degree of dependence and heterogeneity allowable while

<sup>5</sup>There are several other ways of measuring the degree statistical dependence, giving rise to other notions of "mixing." For further details, see Eberlein and Taqqu (1986), Rosenblatt (1956), and White (1984).

still permitting some form of the law of large numbers and the (functional) central limit theorem to obtain. Although (A2) rules out infinite variance marginal distributions of  $\epsilon_t$  such as those in the stable family with characteristic exponent less than 2, the disturbances may still exhibit unconditional leptokurtosis via time-varying conditional moments (e.g., conditional heteroskedasticity). Moreover, since there is a trade-off between (A2) and (A4), the uniform bound on the moments can be relaxed if the mixing coefficients decline faster than (A4) requires.<sup>6</sup> For example, if  $\epsilon_t$  is required to have finite absolute moments of all orders (corresponding to  $\beta \rightarrow \infty$ ), then  $\alpha_k$  must decline faster than  $1/k$ . However, if  $\epsilon_t$  is restricted to have finite moments only up to order 4, then  $\alpha_k$  must decline faster than  $1/k^2$ . These conditions are discussed at greater length by Phillips (1987).

Of course, it is too much to hope that *all* forms of short-memory processes are captured by (A1)–(A4). For example, if  $\epsilon_t$  were the first difference of a stationary process, its spectral density at frequency zero vanishes, violating (A3). Yet such a process certainly need not be long-range dependent. A more subtle example is given by Ibragimov and Rozanov (1978)—a stationary Gaussian process with spectral density function

$$f(\omega) = \exp \left\{ \sum_{k=1}^{\infty} \frac{\cos k\omega}{k \log k + 1} \right\}, \quad (6.2.4)$$

which is strong-mixing but has unbounded spectral density at the origin. The stochastic process with  $1/f(\omega)$  for its spectral density is also strong-mixing, but  $1/f(\omega)$  vanishes at the origin. Although neither process is long-range dependent, they both violate (A3). Unfortunately, a general characterization of the implications of such processes for the behavior of the test statistic proposed in Section 6.3 is currently unavailable. Therefore, a rejection of the null hypothesis does not necessarily imply that long-range dependence is present but merely that, if the rejection is not a type I error, the stochastic process does not satisfy all four conditions simultaneously. Whether or not the composite null (A1)–(A4) is a useful one must therefore depend on the particular application at hand.

In particular, although mixing conditions have been widely used in the recent literature, several other sets of assumptions might have served equally well as our short-range dependent null hypothesis. For example, if  $\{\epsilon_t\}$  is assumed to be stationary and ergodic, the moment condition (A2) can be relaxed and more temporal dependence than (A4) is allowable (see Hall and Heyde, 1980). Whether or not the assumption of stationarity is a

<sup>6</sup>See Herrndorf (1985). One of Mandelbrot's (1972) arguments in favor of *R/S* analysis is that finite second moments are not required. This is indeed the case if we are interested only in the almost sure convergence of the statistic. But since for purposes of inference the *limiting distribution* is required, a stronger moment condition is needed here.

restrictive one for financial time series is still an open question. There is ample evidence of changing variances in stock returns over periods longer than five years, but unstable volatilities can be a symptom of *conditional* heteroskedasticity which can manifest itself in stationary time series. Since the empirical evidence regarding changing conditional moments in asset returns is mixed, allowing for nonstationarities in our null hypothesis may still have value. Moreover, (A1)–(A4) may be weakened further, allowing for still more temporal dependence and heterogeneity, hence widening the class of processes contained in our null hypothesis.<sup>7</sup>

Note, however, that conditions (A1)–(A4) are satisfied by many of the recently proposed stochastic models of persistence, such as those of Campbell and Mankiw (1987), Fama and French (1988), and Poterba and Summers (1988). Therefore, since such models of longer-term correlations are contained in our null, the kind of long-range dependence that (A1)–(A4) were designed to exclude are quite different. Although the distinction between dependence in the short run and the long run may appear to be a matter of degree, strongly dependent processes behave so differently from weakly dependent time series that the dichotomy proposed in our null seems most natural. For example, the spectral densities at frequency zero of strongly dependent processes are either unbounded or zero whereas they are nonzero and finite for processes in our null. The partial sums of strongly dependent processes do not converge in distribution at the same rate as weakly dependent series. And graphically, their behavior is marked by cyclical patterns of all kinds, some that are virtually indistinguishable from trends.

### 6.2.2 Long-Range Dependent Alternatives

In contrast to the short-term memory of “weakly dependent” (i.e., mixing) processes, natural phenomena often display long-term memory in the form of nonperiodic cycles. This has lead several authors to develop stochastic models that exhibit dependence even over very long time spans, such as the fractionally-integrated time series models of Granger (1980), Granger and Joyeux (1980), Hosking (1981), and Mandelbrot and Van Ness (1968). These stochastic processes are not strong-mixing, and have autocorrelation functions that decay at much slower rates than those of weakly dependent processes. For example, let  $X_t$  satisfy the following difference equation:

$$(1 - L)^d X_t = \epsilon_t, \quad \epsilon_t \sim \text{WN}(0, \sigma_\epsilon^2), \quad (6.2.5)$$

where  $L$  is the lag operator and  $\epsilon_t$  is white noise. Granger and Joyeux (1980) and Hosking (1981) show that when the quantity  $(1 - L)^d$  is extended to

<sup>7</sup>Specifically, that the sequence  $\{\epsilon_t\}$  is strong-mixing may be replaced by the weaker assumption that it is a near-epoch dependent function of a strong-mixing process. See McLeish (1977) and Wooldridge and White (1988) for further details.

noninteger powers of  $d$  in the mathematically natural way, the result is a well-defined time series that is said to be “fractionally-differenced” of order  $d$  (or, equivalently, “fractionally-integrated” of order  $-d$ ). Briefly, this involves expanding the expression  $(1 - L)^d$  via the binomial theorem for noninteger powers:

$$(1 - L)^d = \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} L^k, \quad \binom{d}{k} \equiv \frac{d(d-1)(d-2) \cdots (d-k+1)}{k!}, \quad (6.2.6)$$

and then applying the expansion to  $X_t$ :

$$(1 - L)^d X_t = \sum_{k=0}^{\infty} (-1)^k \binom{d}{k} L^k X_t = \sum_{k=0}^{\infty} A_k X_{t-k} = \epsilon_t \quad (6.2.7)$$

where the autoregressive coefficients  $A_k$  are often re-expressed in terms of the gamma function:

$$A_k = (-1)^k \binom{d}{k} = \frac{\Gamma(k-d)}{\Gamma(-d) \Gamma(k+1)}. \quad (6.2.8)$$

$X_t$  may also be viewed mechanically as an infinite-order MA process since

$$X_t = (1 - L)^{-d} \epsilon_t = B(L) \epsilon_t, \quad B_k = \frac{\Gamma(k+d)}{\Gamma(d) \Gamma(k+1)}. \quad (6.2.9)$$

It is not obvious that such a definition of fractional-differencing might yield a useful stochastic process, but Granger (1980), Granger and Joyeux (1980), and Hosking (1981) show that the characteristics of fractionally-differenced time series are interesting indeed. For example, it may be shown that  $X_t$  is stationary and invertible for  $d \in (-\frac{1}{2}, \frac{1}{2})$  (see Hosking, 1981), and exhibits a unique kind of dependence that is positive or negative depending on whether  $d$  is positive or negative, i.e., the autocorrelation coefficients of  $X_t$  are of the same sign as  $d$ . So slowly do the autocorrelations decay that when  $d$  is positive their sum diverges to infinity, and collapses to zero when  $d$  is negative.<sup>8</sup> To develop a sense of this long-range dependence, compare the autocorrelations of a fractionally-differenced  $X_t$  with those of a stationary AR(1) in Table 6.1. Although both the AR(1) and the fractionally-differenced ( $d = \frac{1}{3}$ ) series have first-order autocorrelations of 0.500, at lag 25

<sup>8</sup>Mandelbrot and others have called the  $d < 0$  case “anti-persistence,” reserving the term “long-range dependence” for the  $d > 0$  case. However, since both cases involve autocorrelations that decay much more slowly than those of more conventional time series, we call both long-range dependent.

**Table 6.1.** Comparison of autocorrelation functions of fractionally differenced time series  $(1 - L)^d X_t = \epsilon_t$  for  $d = \frac{1}{3}$ ,  $-\frac{1}{3}$  with that of an AR(1)  $X_t = \rho X_{t-1} + \epsilon_t$ ,  $\rho = .5$ . The variance of  $\epsilon_t$  was chosen to yield a unit variance for  $X_t$  in all three cases.

| Lag $k$ | $\rho(k)$<br>[ $d = \frac{1}{3}$ ] | $\rho(k)$<br>[ $d = -\frac{1}{3}$ ] | $\rho(k)$<br>[AR(1), $\rho = .5$ ] |
|---------|------------------------------------|-------------------------------------|------------------------------------|
| 1       | 0.500                              | -0.250                              | 0.500                              |
| 2       | 0.400                              | -0.071                              | 0.250                              |
| 3       | 0.350                              | -0.036                              | 0.125                              |
| 4       | 0.318                              | -0.022                              | 0.063                              |
| 5       | 0.295                              | -0.015                              | 0.031                              |
| 10      | 0.235                              | -0.005                              | 0.001                              |
| 25      | 0.173                              | -0.001                              | $2.98 \times 10^{-8}$              |
| 50      | 0.137                              | $-3.24 \times 10^{-4}$              | $8.88 \times 10^{-16}$             |
| 100     | 0.109                              | $-1.02 \times 10^{-4}$              | $7.89 \times 10^{-31}$             |

the AR(1) autocorrelation is  $2.98 \times 10^{-8}$  whereas the fractionally-differenced series has autocorrelation 0.173, declining only to 0.109 at lag 100.

In fact, the defining characteristic of long-range dependent processes has been taken by many to be this slow decay of the autocovariance function. Therefore, more generally, long-range dependent processes may be defined to be those processes with autocovariance functions  $\gamma_k$  such that

$$\gamma_k \sim \begin{cases} k^\nu L(k) & \text{for } \nu \in (-1, 0) \text{ or} \\ -k^\nu L(k) & \text{for } \nu \in (-2, -1), \end{cases} \quad \text{as } k \rightarrow \infty, \quad (6.2.10)$$

where  $L(k)$  is any slowly varying function at infinity.<sup>9</sup> This is the definition we shall adopt in the analysis to follow. As an example, the autocovariance function of the fractionally-differenced process (6.2.5) is

$$\gamma_k = \frac{\sigma_\epsilon^2 \Gamma(1 - 2d) \Gamma(k + d)}{\Gamma(d) \Gamma(1 - d) \Gamma(k + 1 - d)} \sim ck^{2d-1} \quad \text{as } k \rightarrow \infty, \quad (6.2.11)$$

where  $d \in (-\frac{1}{2}, \frac{1}{2})$  and  $c$  is some constant. Depending on whether  $d$  is

<sup>9</sup>A function  $f(x)$  is said to be *regularly varying at infinity with index  $\rho$*  if  $\lim_{t \rightarrow \infty} f(tx)/f(t) = x^\rho$  for all  $x > 0$ ; hence regularly varying functions are functions that behave like power functions asymptotically. When  $\rho = 0$ , the function is said to be *slowly varying at infinity*, since it behaves like a constant for large  $x$ . An example of a function that is slowly varying at infinity is  $\log x$ . See Resnick (1987) for further properties of regularly varying functions.

negative or positive, the spectral density of (6.2.5) at frequency zero, given by

$$f(\lambda) \cong (1 - e^{-i\lambda})^{-d}(1 - e^{i\lambda})^{-d}\sigma_\epsilon^2 \sim \sigma_\epsilon^2 \lambda^{-2d} \quad \text{as } \lambda \rightarrow 0, \quad (6.2.12)$$

will either be zero or infinite; thus such processes violate condition (A3).<sup>10</sup> Furthermore, the results of Helson and Sarason (1967) show that these processes are not strong-mixing; hence they also violate condition (A4) of our null hypothesis.<sup>11</sup>

### 6.3 The Rescaled Range Statistic

To detect long-range or “strong” dependence, Mandelbrot has suggested using the range over standard deviation or  $R/S$  statistic, also called the “rescaled range,” which was developed by Hurst (1951) in his studies of river discharges. The  $R/S$  statistic is the range of partial sums of deviations of a time series from its mean, rescaled by its standard deviation. Specifically, consider a sample of returns  $X_1, X_2, \dots, X_n$  and let  $\bar{X}_n$  denote the sample mean  $(1/n) \sum_j X_j$ . Then the classical rescaled range statistic, denoted by  $\tilde{Q}_n$ , is defined as

$$\tilde{Q}_n \equiv \frac{1}{s_n} \left[ \text{Max}_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \bar{X}_n) - \text{Min}_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \bar{X}_n) \right], \quad (6.3.1)$$

where  $s_n$  is the usual (maximum likelihood) standard deviation estimator:

$$s_n \equiv \left[ \frac{1}{n} \sum_j (X_j - \bar{X}_n)^2 \right]^{1/2}. \quad (6.3.2)$$

The first term in brackets in (6.3.1) is the maximum (over  $k$ ) of the partial sums of the first  $k$  deviations of  $X_j$  from the sample mean. Since the sum of all  $n$  deviations of  $X_j$ 's from their mean is zero, this maximum is always nonnegative. The second term in (6.3.1) is the minimum (over  $k$ ) of this same sequence of partial sums; hence it is always nonpositive. The difference of the two quantities, called the “range” for obvious reasons, is always nonnegative, hence  $\tilde{Q}_n \geq 0$ .<sup>12</sup>

<sup>10</sup>This has also been advanced as a definition of long-range dependence—see, for example, Mandelbrot (1972).

<sup>11</sup>Note, Helson and Sarason (1967) only consider the case of linear dependence; hence their conditions are sufficient to rule out strong-mixing but not necessary. For example, white noise may be approximated by a nonlinear deterministic time series (e.g. the tent map) and will have constant spectral density, but will be strongly dependent. We are grateful to Lars Hansen for pointing this out.

<sup>12</sup>The behavior of  $\tilde{Q}_n$  may be better understood by considering its origins in hydrological studies of reservoir design. To accommodate seasonalities in riverflow, a reservoir's capacity

In several seminal papers Mandelbrot, Taqqu, and Wallis demonstrate the superiority of  $R/S$  analysis to more conventional methods of determining long-range dependence, such as analyzing autocorrelations, variance ratios, and spectral decompositions. For example, Mandelbrot and Wallis (1969a) show by Monte Carlo simulation that the  $R/S$  statistic can detect long-range dependence in highly non-Gaussian time series with large skewness and kurtosis. In fact, Mandelbrot (1972, 1975) reports the almost-sure convergence of the  $R/S$  statistic for stochastic processes with infinite variances, a distinct advantage over autocorrelations and variance ratios which need not be well-defined for such processes. Further aspects of the  $R/S$  statistic's robustness are derived in Mandelbrot and Taqqu (1979). Mandelbrot (1972) also argues that, unlike spectral analysis which detects periodic cycles,  $R/S$  analysis can detect *nonperiodic* cycles, cycles with periods equal to or greater than the sample period.

Although these claims may all be contested to some degree, it is a well-established fact that long-range dependence can indeed be detected by the "classical"  $R/S$  statistic. However, perhaps the most important shortcoming of the rescaled range is its sensitivity to short-range dependence, implying that any incompatibility between the data and the predicted behavior of the  $R/S$  statistic under the null hypothesis need not come from long-term memory, but may merely be a symptom of short-term memory.

To see this, first observe that under a simple IID null hypothesis, it is well-known (and is a special case of Theorem 6.3.1 below) that as  $n$  increases without bound, the rescaled range converges in distribution to a well-defined random variable  $V$  when properly normalized, i.e.,

$$\frac{1}{\sqrt{n}} \tilde{Q}_n \Rightarrow V \quad (6.3.3)$$

must be chosen to allow for fluctuations in the supply of water above the dam while still maintaining a relatively constant flow of water below the dam. Since dam construction costs are immense, the importance of estimating the reservoir capacity necessary to meet long term storage needs is apparent. The range is an estimate of this quantity. If  $X_j$  is the riverflow (per unit time) above the dam and  $\bar{X}_n$  is the desired riverflow below the dam, the bracketed quantity in (6.3.1) is the capacity of the reservoir needed to ensure this smooth flow given the pattern of flows in periods 1 through  $n$ . For example, suppose annual riverflows are assumed to be 100, 50, 100, and 50 in years 1 through 4. If a constant annual flow of 75 below the dam is desired each year, a reservoir must have a minimum total capacity of 25 since it must store 25 units in years 1 and 3 to provide for the relatively dry years 2 and 4. Now suppose instead that the natural pattern of riverflow is 100, 100, 50, 50 in years 1 through 4. To ensure a flow of 75 below the dam in this case, the minimum capacity must increase to 50 so as to accommodate the excess storage needed in years 1 and 2 to supply water during the "dry spell" in years 3 and 4. Seen in this context, it is clear that an increase in persistence will increase the required storage capacity as measured by the range. Indeed, it was the apparent persistence of "dry spells" in Egypt that sparked Hurst's life-long fascination with the Nile, leading eventually to his interest in the rescaled range.

Table 6.2. Fractiles of the distribution  $F_V(v)$ .

|            |                        |       |       |       |       |       |       |       |
|------------|------------------------|-------|-------|-------|-------|-------|-------|-------|
| $P(V < v)$ | .005                   | .025  | .050  | .100  | .200  | .300  | .400  | .500  |
| $v$        | 0.721                  | 0.809 | 0.861 | 0.927 | 1.018 | 1.090 | 1.157 | 1.223 |
| $P(V < v)$ | .543                   | .600  | .700  | .800  | .900  | .950  | .975  | .995  |
| $v$        | $\sqrt{\frac{\pi}{2}}$ | 1.294 | 1.374 | 1.473 | 1.620 | 1.747 | 1.862 | 2.098 |

where “ $\Rightarrow$ ” denotes weak convergence and  $V$  is the range of a Brownian bridge on the unit interval.<sup>13</sup>

Now suppose, instead, that  $\{X_j\}$  were short-range dependent—for example, let  $X_j$  be a stationary AR(1):<sup>14</sup>

$$\epsilon_t = \rho\epsilon_{t-1} + \eta_t, \quad \eta_t \sim \text{WN}(0, \sigma_\eta^2), \quad |\rho| \in (0, 1). \quad (6.3.4)$$

Although  $\{\epsilon_t\}$  is short-range dependent, it yields a  $\tilde{Q}_n$  that does not satisfy (6.3.3). In fact, it may readily be shown that for (6.3.4) the limiting distribution of  $\tilde{Q}_n/\sqrt{n}$  is  $\xi V$  where  $\xi \equiv \sqrt{(1+\rho)/(1-\rho)}$  (see Proposition 6.3.1 below). For some portfolios of common stock,  $\hat{\rho}$  is as large as 50 percent, implying that the mean of  $\tilde{Q}_n/\sqrt{n}$  may be biased upward by 73 percent! Since the mean of  $V$  is  $\sqrt{\pi/2} \approx 1.25$ , the mean of the classical rescaled range would be 2.16 for such an AR(1) process. Using the critical values of  $V$  reported in Table 6.2, it is evident that a value of 2.16 would yield a rejection of the null hypothesis at any conventional significance level.

This should come as no surprise since the values in Table 6.2 correspond to the distribution of  $V$ , not  $\xi V$ . Now by taking into account the “short-term” autocorrelations of the  $X_j$ ’s—by dividing  $Q_n$  by  $\xi$  for example—convergence to  $V$  may be restored. But this requires knowledge of  $\xi$  which, in turn, requires knowledge of  $\rho$ . Moreover, if  $X_j$  follows a short-range dependent process other than an AR(1), the expression for  $\xi$  will change, as Proposition 6.3.1 below shows. Therefore, correcting for short-range dependence on a case-by-case basis is impractical. Ideally, we would like to correct for short-term memory without taking too strong a position on what form it takes. This is precisely what the modified rescaled range of Section 6.3.1 does—its limiting distribution is invariant to many forms of short-range dependence, and yet it is still sensitive to the presence of long-range dependence.

<sup>13</sup>See Billingsley (1968) for the definition of weak convergence. We discuss the Brownian bridge and  $V$  more formally below.

<sup>14</sup>It is implicitly assumed throughout that white noise has a Lebesgue-integrable characteristic function to avoid the pathologies of Andrews (1984).

Although aware of the effects of short-range dependence on the rescaled range, Mandelbrot (1972, 1975) did not correct for this bias since his focus was the relation of the  $R/S$  statistic's logarithm to the logarithm of the sample size as the sample size increases without bound. For short-range dependent time series such as strong-mixing processes the ratio  $\log \tilde{Q}_n / \log n$  approaches  $\frac{1}{2}$  in the limit, but converges to quantities greater or less than  $\frac{1}{2}$  according to whether there is positive or negative long-range dependence. The limit of this ratio is often denoted by  $H$  and is called the "Hurst" coefficient. For example, the fractionally-differenced process (6.2.1) satisfies the simple relation:  $H = d + \frac{1}{2}$ .

Mandelbrot and Wallis (1969a) suggest estimating the Hurst coefficient by plotting the logarithm of  $\tilde{Q}_n$  against the logarithm of the sample size  $n$ . Beyond some large  $n$ , the slope of such a plot should settle down to  $H$ . However, although  $H = \frac{1}{2}$  across general classes of short-range dependent processes, the finite-sample properties of the *estimated* Hurst coefficient are not invariant to the form of short-range dependence. In particular, Davies and Harte (1987) show that even though the Hurst coefficient of a stationary Gaussian AR(1) is precisely  $\frac{1}{2}$ , the 5 percent Mandelbrot regression test rejects this null hypothesis 47 percent of the time for an autoregressive parameter of 0.3. Additional Monte Carlo evidence is reported in Section 6.5.

### 6.3.1 The Modified $R/S$ Statistic

To distinguish between long-range and short-range dependence, the  $R/S$  statistic must be modified so that its statistical behavior is invariant over a general class of short memory processes, but deviates for long memory processes. This is accomplished by the following statistic  $Q_n$ :

$$Q_n \equiv \frac{1}{\hat{\sigma}_n(q)} \left[ \text{Max}_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \bar{X}_n) - \text{Min}_{1 \leq k \leq n} \sum_{j=1}^k (X_j - \bar{X}_n) \right] \quad (6.3.5)$$

where

$$\begin{aligned} \hat{\sigma}_n^2(q) &\equiv \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2 \\ &\quad + \frac{2}{n} \sum_{j=1}^q \omega_j(q) \left\{ \sum_{i=j+1}^n (X_i - \bar{X}_n)(X_{i-j} - \bar{X}_n) \right\} \end{aligned} \quad (6.3.6)$$

$$= \hat{\sigma}_x^2 + 2 \sum_{j=1}^q \omega_j(q) \hat{\gamma}_j, \quad \omega_j(q) \equiv 1 - \frac{j}{q+1}, \quad q < n, \quad (6.3.7)$$

and  $\hat{\sigma}_x^2$  and  $\hat{\gamma}_j$  are the usual sample variance and autocovariance estimators of  $X$ .

$Q_n$  differs from  $\tilde{Q}_n$  only in its denominator, which is the square root of a consistent estimator of the partial sum's variance. If  $\{X_t\}$  is subject to short-range dependence, the variance of the partial sum is not simply the sum of the variances of the individual terms, but also includes the autocovariances. Therefore, the estimator  $\hat{\sigma}_n(q)$  involves not only sums of squared deviations of  $X_j$ , but also its weighted autocovariances up to lag  $q$ . The weights  $\omega_j(q)$  are those suggested by Newey and West (1987) and always yield a positive  $\hat{\sigma}_n^2(q)$ , an estimator of  $2\pi$  times the (unnormalized) spectral density function of  $X_t$  at frequency zero using a Bartlett window. Theorem 4.2 of Phillips (1987) demonstrates the consistency of  $\hat{\sigma}_n(q)$  under the following conditions:<sup>15</sup>

$$(A2') \quad \sup_t E[|\epsilon_t|^{2\beta}] < \infty \text{ for some } \beta > 2.$$

(A5) As  $n$  increases without bound,  $q$  also increases without bound such that  $q \sim o(n^{1/4})$ .

By allowing  $q$  to increase with (but at a slower rate than) the number of observations  $n$ , the denominator of  $Q_n$  adjusts appropriately for general forms of short-range dependence. Of course, although the conditions (A2') and (A5) ensure the consistency of  $\hat{\sigma}^2(q)$ , they provide little guidance in selecting a truncation lag  $q$ . Monte Carlo studies such as Andrews (1991) and Lo and MacKinlay (1989a) have shown that when  $q$  becomes large relative to the sample size  $n$ , the finite-sample distribution of the estimator can be radically different from its asymptotic limit. However  $q$  cannot be chosen too small since the autocovariances beyond lag  $q$  may be substantial and should be included in the weighted sum. Therefore, the truncation lag must be chosen with some consideration of the data at hand. Andrews (1991) does provide a data-dependent rule for choosing  $q$ , however its minimax optimality is still based on an asymptotic mean-squared error criterion—little is known about how best to pick  $q$  in finite samples. Some Monte Carlo evidence is reported in Section 6.5.

Since there are several other consistent estimators of the spectral density function at frequency zero, conditions (A2') and (A5) can be replaced with weaker assumptions if conditions (A1), (A3), and (A4) are suitably modified. If, for example,  $X_t$  is  $m$ -dependent (so that observations spaced greater than  $m$  periods apart are independent), it is well-known that the spectral density at frequency zero may be estimated consistently with a finite number of unweighted autocovariances (see, for example, Hansen, 1982, Lemma 3.2). Other weighting functions may be found in Hannan (1970, Chapter V.4) and may yield better finite-sample properties for  $Q_n$  than the Bartlett

<sup>15</sup> Andrews (1991) has improved the rate restriction in (A5) to  $o(n^{1/2})$ , and it has been conjectured that  $o(n)$  is sufficient.

window without altering the limiting null distribution derived in the next section.<sup>16</sup>

### 6.3.2 The Asymptotic Distribution of $Q_n$

To derive the limiting distribution of the modified rescaled range  $Q_n$  under our null hypothesis, consider the behavior of the following standardized partial sum:

$$W_n(\tau) \equiv \frac{1}{\sigma\sqrt{n}} S_{[n\tau]}, \quad \tau \in [0, 1], \quad (6.3.8)$$

where  $S_k$  denotes the partial sum  $\sum_{j=1}^k \epsilon_j$  and  $[n\tau]$  is the greatest integer less than or equal to  $n\tau$ . The sample paths of  $W_n(\tau)$  are elements of the function space  $\mathcal{D}[0, 1]$ , the space of all real-valued functions on  $[0, 1]$  that are right-continuous and possess finite left limits. Under certain conditions it may be shown that  $W_n(\tau)$  converges weakly to a Brownian motion  $W(\tau)$  on the unit interval, and that well-behaved functionals of  $W_n(\tau)$  converge weakly to the same functionals of Brownian motion (see Billingsley, 1968, for further details). Armed with these results, the limiting distribution of the modified rescaled range may be derived in three easy steps, summarized in the following theorem.<sup>17</sup>

**Theorem 6.3.1.**<sup>18</sup> *If  $\{\epsilon_t\}$  satisfies assumptions (A1), (A2'), (A3)–(A5), then as  $n$  increases without bound:*

- (a)  $\text{Max}_{1 \leq k \leq n} \frac{1}{\hat{\sigma}_n(q)\sqrt{n}} \sum_{j=1}^k (X_j - \bar{X}_n) \Rightarrow \text{Max}_{0 \leq \tau \leq 1} W^\circ(\tau) \equiv M^\circ,$
- (b)  $\text{Min}_{1 \leq k \leq n} \frac{1}{\hat{\sigma}_n(q)\sqrt{n}} \sum_{j=1}^k (X_j - \bar{X}_n) \Rightarrow \text{Min}_{0 \leq \tau \leq 1} W^\circ(\tau) \equiv m^\circ,$
- (c)  $\frac{1}{\sqrt{n}} Q_n \Rightarrow M^\circ - m^\circ \equiv V.$

Parts (a) and (b) of Theorem 6.3.1 follow from Lemmas A.1 and A.2 of the Appendix, and Theorem 4.2 of Phillips (1987), and show that the maximum and minimum of the partial sum of deviations of  $X_j$  from its mean converge respectively to the maximum and minimum of the celebrated Brownian bridge  $W^\circ(\tau)$  on the unit interval, also called “pinned” or “tied-down” Brownian motion because  $W^\circ(0) = W^\circ(1) = 0$ . That the limit of

<sup>16</sup>For example, Andrews (1991) and Gallant (1987) both advocate the use of Parzen weights, which also yields a positive semi-definite estimator of the spectral density at frequency zero but is optimal in an asymptotic mean-squared error sense.

<sup>17</sup>Mandelbrot (1975) derives similar limit theorems for the statistic  $\tilde{Q}_n$  under the more restrictive IID assumption, in which case the limiting distribution will coincide with that of  $Q_n$ . Since our null hypothesis includes weakly dependent disturbances, we extend his results via the more general functional central limit theorem of Herrndorf (1984, 1985).

<sup>18</sup>Proofs of theorems are given in the Appendix.

the partial sums is a Brownian bridge is not surprising since the summands are deviations from the mean and must therefore sum to zero at  $k = n$ . Part (6.3.1) of the theorem follows immediately from Lemma A.2 and is the key result, allowing us to perform large sample statistical inference once the distribution function for the range of the Brownian bridge is obtained. This distribution function is implicitly contained in Feller (1951), and is given explicitly by Kennedy (1976) and Siddiqui (1976) as<sup>19</sup>

$$F_V(v) = 1 + 2 \sum_{k=1}^{\infty} (1 - 4k^2 v^2) e^{-(kv)^2}. \quad (6.3.9)$$

Critical values for tests of any significance level are easily obtained from this simple expression (6.3.9) for  $F_V$ . The values most commonly used are reported in Table 6.2. The moments of  $V$  may also be readily computed from (6.3.9); a simple calculation shows that  $E[V] = \sqrt{\pi/2}$  and  $E[V^2] = \pi^2/6$ , thus the mean and standard deviation of  $V$  are approximately 1.25 and 0.27 respectively. Plots of  $F_V$  and  $f_V$  are given in Figure 6.1, along with Gaussian distribution and density functions (with the same mean and variance as  $V$ ) for comparison. The distribution of  $V$  is positively skewed and most of its mass falls between  $\frac{3}{4}$  and 2.

### 6.3.3 The Relation Between $Q_n$ and $\tilde{Q}_n$

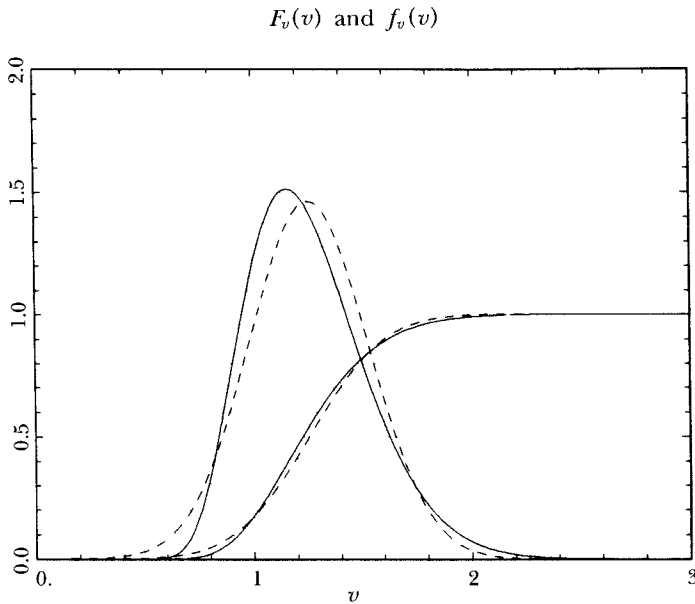
Since  $Q_n$  and  $\tilde{Q}_n$  differ solely in how the range is normalized, the limiting behavior of our modified  $R/S$  statistic and Mandelbrot's original will only coincide when  $\hat{\sigma}_n(q)$  and  $s_n$  are asymptotically equivalent. From the definitions of  $\hat{\sigma}_n(q)$  and  $s_n$ , it is apparent that the two will generally converge in probability to different limits in the presence of autocorrelation. Therefore, under the weakly dependent null hypothesis the statistic  $\tilde{Q}_n/\sqrt{n}$  will converge to the range  $V$  of a Brownian bridge multiplied by some constant. More formally, we have the almost trivial result:

**Proposition 6.3.1.** *If  $\lim_{n \rightarrow \infty} E[\sum_{j=1}^n \epsilon_j^2/n]$  is finite and positive, then under assumptions (A1)–(A4),  $\tilde{Q}_n/\sqrt{n} \Rightarrow \xi V$  where*

$$\xi^2 \equiv \frac{\lim_{n \rightarrow \infty} E\left[\frac{1}{n} \left(\sum_{j=1}^n \epsilon_j\right)^2\right]}{\lim_{n \rightarrow \infty} E\left[\frac{1}{n} \sum_{j=1}^n \epsilon_j^2\right]}. \quad (6.3.10)$$

Therefore, normalizing the range by  $s_n$  in place of  $\hat{\sigma}_n(q)$  changes the limiting distribution of the rescaled range by the multiplicative constant

<sup>19</sup>We are grateful to David Aldous and Yin-Wong Cheung for these last two references.



**Figure 6.1.** Distribution and density function of the range  $V$  of a Brownian bridge. Dashed curves are the normal distribution and density functions with mean and variance equal to those of  $V$  ( $\sqrt{\pi/2}$  and  $\pi^2/6$  respectively).

$\xi$ . This result was used above to derive the limiting distribution of  $\tilde{Q}_n$  in the AR(1) case, and closed-form expressions for  $\xi$  for general stationary ARMA( $p, q$ ) processes may readily be obtained using (6.3.10).

Since it is robust to many forms of heterogeneity and weak dependence, tests based on the modified  $R/S$  statistic  $Q_n$  cover a broader set of null hypotheses than those using  $\tilde{Q}_n$ . More to the point, the modified rescaled range is able to distinguish between short-range and long-range dependence—the classical rescaled range cannot. Whereas an extreme value for  $Q_n$  indicates the likelihood of long-range dependence, a rejection based on the  $\tilde{Q}_n$  statistic is also consistent with short-range dependence in the data. Of course, it is always possible to tabulate the limiting distribution of the classical  $R/S$  statistic under a particular model of short-range dependence, but this obviously suffers from the drawback of specificity. The modified rescaled range converges weakly to the range of a Brownian bridge under *general* forms of weak dependence.

Despite its sensitivity to short-range dependence, the classical  $R/S$  statistic may still be used to test for independently and identically distributed  $X_t$ 's. Indeed, the AR(1) example of Section 6.3 and the results of Davies and

Harte (1987) suggest that such a test may have considerable power against non-IID alternatives. However, since there is already a growing consensus among financial economists that stock market prices are not independently and identically distributed, this null hypothesis is of less immediate interest. For example, it is now well-known that aggregate stock market returns exhibit significant serial dependence for short-horizon holding periods and are therefore not independently distributed.

#### 6.3.4 The Behavior of $Q_n$ Under Long Memory Alternatives

To complete the analysis of the modified rescaled range, its behavior under long-range dependent alternatives remains to be investigated. Although this depends of course on the specific alternative at hand, surprisingly general results are available based on the following result from Taqqu (1975).

**Theorem 6.3.2 (Taqqu).** *Let  $\{\epsilon_i\}$  be a zero-mean stationary Gaussian stochastic process such that:*

$$\sigma_n^2 \equiv \text{Var}[S_n] \sim n^{2H} L(n) \quad (6.3.11)$$

where  $S_n$  is the partial sum  $\sum_{j=1}^n \epsilon_j$ ,  $H \in (0, 1)$ , and  $L(n)$  is a slowly varying function at infinity. Define the following function on  $\mathcal{D}[0, 1]$ :

$$W_n(\tau) \equiv \frac{1}{\sigma_n} S_{[n\tau]}, \quad \tau \in (0, 1). \quad (6.3.12)$$

Then  $W_n(\tau) \Rightarrow W_H(\tau)$ , where  $W_H(\tau)$  is a fractional Brownian motion of order  $H$  on  $[0, 1]$ .

Theorem 6.3.2 is a functional central limit theorem for strongly dependent processes, and is only a special case of Taqqu's (1975) considerably more general results. In contrast to the usual functional central limit theorem in which properly normalized partial sums converge to a standard Brownian motion, Theorem 6.3.2 states that long-range dependent partial sums converge weakly to a fractional Brownian motion, first defined by Mandelbrot and Van Ness (1968) as the following stochastic integral:

$$W_H(\tau) \equiv \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^\tau (\tau - x)^{H-\frac{1}{2}} dW(x). \quad (6.3.13)$$

Observe that when  $H = \frac{1}{2}$ ,  $W_H(\tau)$  reduces to a standard Brownian motion. In that case, there is no long-range dependence, the variance of the partial sums grows at rate  $n$ , and the spectral density at frequency zero is finite and positive. If  $H \in (\frac{1}{2}, 1)$  ( $H \in (0, \frac{1}{2})$ ), there is positive (negative) long-range dependence, the variance grows faster (slower) than  $n$ , hence the spectral density at frequency zero is infinite (zero).

In a fashion analogous to Theorem 6.3.1, the behavior of  $Q_n$  under long-range dependent alternatives may now be derived in several steps using Lemmas A.2, A.3, and Theorem 6.3.2:

**Theorem 6.3.3.** *Let  $\{\epsilon_t\}$  be a zero-mean stationary Gaussian stochastic process with autocovariance function  $\gamma_k$  such that*

$$\gamma_k \sim \begin{cases} k^{2H-2} L(k) & \text{for } H \in (\frac{1}{2}, 1) \text{ or,} \\ -k^{2H-2} L(k) & \text{for } H \in (0, \frac{1}{2}) \end{cases} \quad \text{as } k \rightarrow \infty \quad (6.3.14)$$

where  $L(k)$  is a slowly varying function at infinity. Then as  $n$  and  $q$  increase without bound such that  $(q/n) \rightarrow 0$ , we have:

$$(a) \quad \text{Max}_{1 \leq k \leq n} \frac{1}{\sigma_n} \sum_{j=1}^k (X_j - \bar{X}_n) \Rightarrow \text{Max}_{0 \leq \tau \leq 1} W_H^\circ(\tau) \equiv M_H^\circ,$$

$$(b) \quad \text{Min}_{1 \leq k \leq n} \frac{1}{\sigma_n} \sum_{j=1}^k (X_j - \bar{X}_n) \Rightarrow \text{Min}_{0 \leq \tau \leq 1} W_H^\circ(\tau) \equiv m_H^\circ,$$

$$(c) \quad R_n \equiv \frac{\hat{\sigma}_n(q)\sqrt{n}}{\sigma_n} \cdot \frac{1}{\sqrt{n}} Q_n \Rightarrow M_H^\circ - m_H^\circ \equiv V_H,$$

$$(d) \quad a_n \equiv \frac{\sigma_n}{\hat{\sigma}_n(q)\sqrt{n}} \xrightarrow{p} \begin{cases} \infty & \text{for } H \in (\frac{1}{2}, 1), \\ 0 & \text{for } H \in (0, \frac{1}{2}), \end{cases}$$

$$(e) \quad \frac{1}{\sqrt{n}} Q_n = a_n R_n \xrightarrow{p} \begin{cases} \infty & \text{for } H \in (\frac{1}{2}, 1), \\ 0 & \text{for } H \in (0, \frac{1}{2}), \end{cases}$$

where  $\hat{\sigma}_n(q)$  is defined in (6.3.6),  $\sigma_n$  is defined in Theorem 6.3.2, and  $W_H^\circ(\tau) \equiv W_H(\tau) - \tau W_H(1)$ .<sup>20</sup>

Theorem 6.3.3 shows that the modified rescaled range test is consistent against a class of long-range dependent stationary Gaussian alternatives. In the presence of positive strong dependence, the  $R/S$  statistic diverges in probability to infinity; in the presence of negative strong dependence, it converges in probability to zero. In either case, the probability of rejecting the null hypothesis approaches unity for all stationary Gaussian stochastic processes satisfying (6.3.14), a broad set of alternatives that includes all fractionally-differenced Gaussian ARIMA( $p, d, q$ ) models with  $d \in (-\frac{1}{2}, \frac{1}{2})$ .

From (a) and (b) of Theorem 6.3.3 it is apparent that the normalized population rescaled,  $R_n/\sqrt{n}$ , converges to zero in probability. Therefore,

<sup>20</sup>Although it is tempting to call  $W_H^\circ(\tau)$  a "fractional Brownian bridge," this is not the most natural definition despite the fact that it is "tied down." See Jonas (1983, Chapter 3.3) for a discussion.

whether or not  $Q_n/\sqrt{n}$  approaches zero or infinity in the limit depends entirely on the limiting behavior of the ratio  $\sigma_n/\hat{\sigma}_n(q)$ . That is,

$$\frac{Q_n}{\sqrt{n}} = \frac{\sigma_n}{\hat{\sigma}_n(q)} \frac{R_n}{\sqrt{n}} \quad (6.3.15)$$

so that if the ratio  $\sigma_n/\hat{\sigma}_n(q)$  diverges fast enough to overcompensate for the convergence of  $R_n/\sqrt{n}$  to zero, then the test will reject in the upper tail, otherwise it will reject in the lower tail. This is determined by whether  $d$  lies in the interval  $(0, \frac{1}{2})$  or  $(-\frac{1}{2}, 0)$ . When  $d = 0$ , the ratio  $\sigma_n/\hat{\sigma}_n(q)$  converges to unity in probability and, as expected, the normalized  $R/S$  statistic converges in distribution to the range of the standard Brownian bridge.

Of course, if one is interested exclusively in fractionally-differenced alternatives, a more efficient means of detecting long-range dependence might be to estimate the fractional differencing parameter directly. In such cases, the approaches taken by Geweke and Porter-Hudak (1983), Sowell (1990), and Yajima (1985, 1988) may be preferable. The modified  $R/S$  test is perhaps most useful for detecting departures into a broader class of alternative hypotheses, a kind of “portmanteau” test statistic that may complement a comprehensive analysis of long-range dependence.

## 6.4 *R/S Analysis for Stock Market Returns*

The importance of long-range dependence in asset markets was first considered by Mandelbrot (1971). More recently, the evidence uncovered by Fama and French (1988), Lo and MacKinlay (1988b), and Poterba and Summers (1988) may be symptomatic of a long-range dependent component in stock market prices. In particular, Lo and MacKinlay (1988b) show that the ratios of  $k$ -week stock return variances to  $k$  times the variance of one-week returns generally exceed unity when  $k$  is small (2 to 32). In contrast, Poterba and Summers (1988) find that this same variance ratio falls below one when  $k$  is much larger (96 and greater).

To see that such a phenomenon can easily be generated by long-range dependence, denote by  $X_t$  the time- $t$  return on a stock and let it be the sum of two components  $X_{at}$  and  $X_{bt}$  where

$$(1 - L)^d X_{at} = \epsilon_t, \quad (1 - \rho L) X_{bt} = \eta_t, \quad (6.4.1)$$

and assign the values  $(-0.2, 0.25, 1, 1.1)$  to the parameters  $(d, \rho, \sigma_\epsilon^2, \sigma_\eta^2)$ . Let the ratio of the  $k$ -period return variance to  $k$  times the variance of  $X_t$

be denoted by  $VR(k)$ . Then a simple calculation will show that for the parameter values chosen:

$$\begin{array}{ll} VR(2) = 1.04, & VR(10) = 10.4, \\ VR(3) = 1.06, & VR(50) = 0.97, \\ VR(4) = 1.07, & VR(100) = 0.95, \\ VR(5) = 1.06, & VR(250) = 0.92. \end{array}$$

The intuition for this pattern of variance ratios comes from observing that  $VR(k)$  is a weighted sum of the first  $k - 1$  autocorrelation coefficients of  $X_t$  with linearly declining weights (see Lo and MacKinlay, 1988b). When  $k$  is small the autocorrelation of  $X_t$  is dominated by the positively autocorrelated  $AR(1)$  component  $X_{bt}$ . But since the autocorrelations of  $X_{bt}$  decay rapidly relative to those of  $X_{at}$ , as  $k$  grows the influence of the long-memory component eventually outweighs that of the  $AR(1)$ , ultimately driving the variance ratio below unity.

#### 6.4.1 The Evidence for Weekly and Monthly Returns

Greene and Fielitz (1977) were perhaps the first to apply  $R/S$  analysis to common stock returns. More recent applications include Booth and Kaen (1979) (gold prices), Booth, Kaen, and Koveos (1982) (foreign exchange rates), and Helms, Kaen, and Rosenman (1984) (futures contracts). These and earlier applications of  $R/S$  analysis by Mandelbrot and Wallis (1969a) have three features in common: (1) They provide no sampling theory with which to judge the statistical significance of their empirical results; (2) they use the  $\bar{Q}_n$  statistic which is not robust to short-range dependence; and (3) they do not focus on the  $R/S$  statistic itself, but rather on the regression of its logarithm on (sub)sample sizes. The shortcomings of (1) and (2) are apparent from the discussion in the preceding sections. As for (3), Davies and Harte (1987) show such regression tests to be significantly biased toward rejection even for a stationary  $AR(1)$  process with an autoregressive parameter of 0.3.

To test for long-term memory in stock returns, we use data from the Center for Research in Security Prices (CRSP) monthly and daily returns files. Tests are performed for the value- and equal-weighted CRSP indexes. Daily observations for the returns indexes are available from July 3, 1962, to December 31, 1987 yielding a sample size of 6,409 observations. Monthly indexes are each composed of 744 observations from January 30, 1926, to December 31, 1987. The following statistic is computed for the various returns indexes:

$$V_n(q) \equiv \frac{1}{\sqrt{n}} Q_n \stackrel{a}{\sim} V, \quad (6.4.2)$$

where the distribution  $F_V$  of  $V$  is given in (6.3.9). Using the values in Table 6.2 a test of the null hypothesis may be performed at the 95 percent level

of confidence by accepting or rejecting according to whether  $V_n$  is or is not contained in the interval  $[0.809, 1.862]$  which assigns equal probability to each tail.

$V_n(q)$  is written as a function of  $q$  to emphasize the dependence of the modified rescaled range on the truncation lag. To check the sensitivity of the statistic to the lag length,  $V_n(q)$  is computed for several different values of  $q$ . The normalized classical Hurst-Mandelbrot rescaled range  $\tilde{V}_n$  is also computed for comparison, where

$$\tilde{V}_n \equiv \frac{1}{\sqrt{n}} \tilde{Q}_n \stackrel{a}{\sim} \xi V. \quad (6.4.3)$$

Table 6.3 reports results for the daily equal- and value-weighted returns indexes. The panel labelled “Equal-Weighted” contains the  $V_n(q)$  and  $\tilde{V}_n$  statistics for the equal-weighted index for the entire sample period (the first row), two equally-partitioned sub-samples (the next two rows), and four equally-partitioned sub-samples (the next four rows). The modified rescaled range is computed with  $q$ -values of 90, 180, 270, and 360 days. The columns labelled “%-Bias” report the estimated bias of the original rescaled range  $\tilde{V}_n$ , and is  $100 \cdot (\hat{\xi} - 1)$  where  $\hat{\xi} = \hat{\sigma}_n(q)/s_n = \tilde{V}_n/V_n$ .

Although Table 6.3 shows that the classical  $R/S$  statistic  $\tilde{V}_n$  is statistically significant at the 5 percent level for the daily equal-weighted CRSP returns index, the modified  $R/S$  statistic  $V_n$  is not. While  $\tilde{V}_n$  is 2.63 for the entire sample period the modified  $R/S$  statistic is 1.46 with a truncation lag of 90 days, and 1.50 with a truncation lag of 360 days. The importance of normalizing by  $\hat{\sigma}_n(q)$  is clear—dividing by  $s_n$  imparts a potential upward bias of 80 percent!

The statistical insignificance of the modified  $R/S$  statistics indicates that the data are consistent with the short-memory null hypothesis. The stability of the  $V_n(q)$  across truncation lags  $q$  also supports the hypothesis that there is little dependence in daily stock returns beyond one or two months. For example, using 90 lags yields a  $V_n$  of 1.46 whereas 270 and 360 lags both yield 1.50, virtually the same point estimate. The results are robust to the sample period—none of the sub-period  $V_n(q)$ ’s are significant. The classical rescaled range is significant only in the first half of the sample for the value-weighted index, and is insignificant when the entire sample is used.

Table 6.4 reports similar results for monthly returns indexes with four values of  $q$  employed: 3, 6, 9, and 12 months. None of the modified  $R/S$  statistics are statistically significant at the 5 percent level in any sample period or sub-period for either index. The percentage bias is generally lower for monthly data, although it still ranges from  $-0.2$  to 25.3 percent.

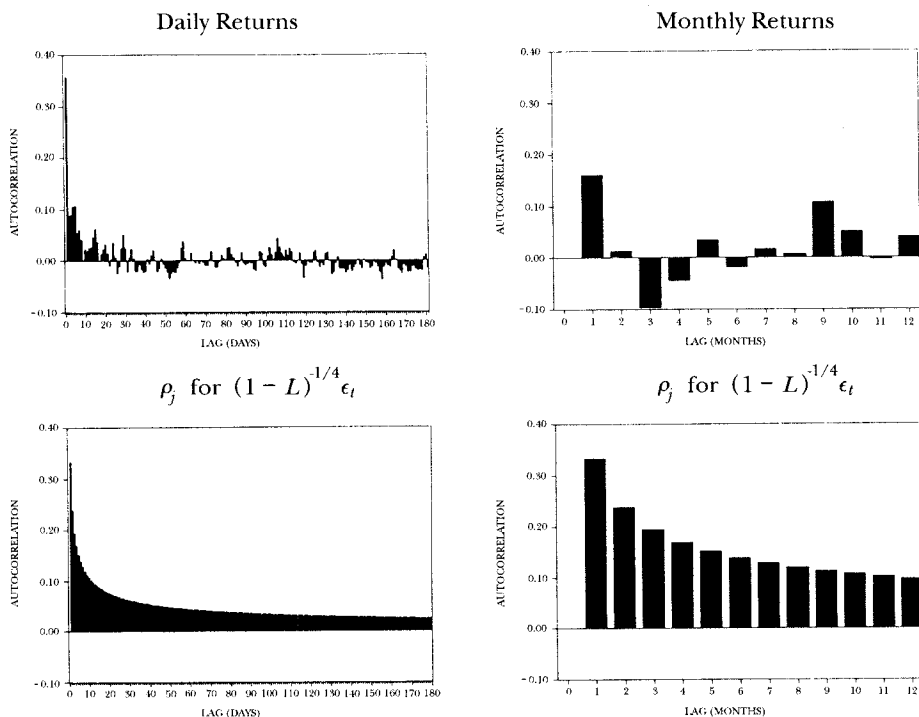
To develop further intuition for these results, Figure 6.2 contains the autocorrelograms of the daily and monthly equal-weighted returns indexes, where the maximum lag is 360 for daily returns and 12 for monthly. For both

**Table 6.3.** *R/S analysis of daily equal- and value-weighted CRSP stock returns indexes from July 3, 1962, to December 31, 1987 using the classical rescaled range  $\tilde{V}_n$  and the modified rescaled range  $V_n(q)$ . Entries in the %-bias columns are computed as  $[(\tilde{V}_n/V_n(q)) - 1] \cdot 100$ , and are estimates of the bias of the classical R/S statistic in the presence of short-term dependence. Asterisks indicate significance at the 5 percent level.*

| Time Period            | Sample Size | $\tilde{V}_n$ | $V_n(90)$ | %-Bias | $V_n(180)$ | %-Bias | $V_n(270)$ | %-Bias | $V_n(360)$ | %-Bias |
|------------------------|-------------|---------------|-----------|--------|------------|--------|------------|--------|------------|--------|
| <i>Equal-Weighted:</i> |             |               |           |        |            |        |            |        |            |        |
| 620703–871231          | 6409        | 2.63*         | 1.46      | 79.9   | 1.45       | 81.1   | 1.50       | 75.2   | 1.50       | 75.4   |
| 620703–750428          | 3204        | 3.18*         | 1.61      | 97.0   | 1.57       | 102.0  | 1.63       | 95.2   | 1.62       | 96.8   |
| 750429–871231          | 3205        | 1.45          | 0.92      | 57.2   | 0.97       | 49.0   | 1.05       | 38.5   | 1.14       | 27.3   |
| 620703–681217          | 1602        | 2.40*         | 1.39      | 72.2   | 1.46       | 64.7   | 1.72       | 39.7   | 1.78       | 34.8   |
| 681219–750428          | 1602        | 2.03*         | 1.07      | 90.7   | 1.10       | 84.9   | 1.19       | 70.6   | 1.23       | 65.3   |
| 750428–810828          | 1602        | 1.35          | 0.89      | 51.6   | 1.23       | 9.5    | 1.49       | –9.2   | 1.71       | –21.0  |
| 810831–871231          | 1603        | 1.79          | 1.15      | 55.8   | 1.10       | 62.4   | 1.18       | 51.6   | 1.27       | 41.4   |
| <i>Value-Weighted:</i> |             |               |           |        |            |        |            |        |            |        |
| 620703–871231          | 6409        | 1.55          | 1.29      | 20.8   | 1.26       | 22.9   | 1.30       | 19.1   | 1.33       | 16.8   |
| 620703–750428          | 3204        | 1.97*         | 1.43      | 37.3   | 1.39       | 41.4   | 1.43       | 37.5   | 1.45       | 35.5   |
| 750429–871231          | 3205        | 1.29          | 1.22      | 5.8    | 1.24       | 4.1    | 1.32       | –2.3   | 1.42       | –9.4   |
| 620703–681217          | 1602        | 1.67          | 1.43      | 16.8   | 1.45       | 15.3   | 1.62       | 3.4    | 1.69       | –1.3   |
| 681219–750428          | 1602        | 1.85          | 1.34      | 38.2   | 1.34       | 38.2   | 1.40       | 31.7   | 1.45       | 27.1   |
| 750428–810828          | 1602        | 1.08          | 1.12      | –3.7   | 1.26       | –14.7  | 1.34       | –19.4  | 1.42       | –24.2  |
| 810831–871231          | 1603        | 1.50          | 1.38      | 8.8    | 1.37       | 9.2    | 1.50       | –0.3   | 1.63       | –8.0   |

**Table 6.4.** *R/S analysis of monthly equal- and value-weighted CRSP stock returns indexes from January 30, 1926, to December 31, 1987, using the classical rescaled range  $\tilde{V}_n$  and the modified rescaled range  $V_n(q)$ . Entries in the %-bias columns are computed as  $[(\tilde{V}_n/V_n(q)) - 1] \cdot 100$ , and are estimates of the bias of the classical R/S statistic in the presence of short-term dependence. Asterisks indicate significance at the 5 percent level.*

| Time Period            | Sample Size | $\tilde{V}_n$ | $V_n(3)$ | %-Bias | $V_n(6)$ | %-Bias | $V_n(9)$ | %-Bias | $V_n(12)$ | %-Bias |
|------------------------|-------------|---------------|----------|--------|----------|--------|----------|--------|-----------|--------|
| <i>Equal-Weighted:</i> |             |               |          |        |          |        |          |        |           |        |
| 260130–871231          | 744         | 1.17          | 1.07     | 9.1    | 1.10     | 6.6    | 1.09     | 7.2    | 1.06      | 10.4   |
| 260130–561231          | 372         | 1.32          | 1.21     | 9.4    | 1.26     | 5.1    | 1.24     | 7.1    | 1.18      | 12.1   |
| 570131–871231          | 372         | 1.37          | 1.26     | 8.4    | 1.23     | 11.1   | 1.27     | 7.6    | 1.30      | 5.2    |
| 260130–410630          | 186         | 1.42          | 1.31     | 8.3    | 1.40     | 1.6    | 1.39     | 2.6    | 1.32      | 8.0    |
| 410731–561231          | 186         | 1.60          | 1.42     | 13.1   | 1.34     | 20.0   | 1.28     | 25.3   | 1.28      | 25.1   |
| 570131–720630          | 186         | 1.20          | 1.04     | 15.9   | 0.99     | 21.9   | 1.03     | 17.4   | 1.07      | 12.3   |
| 720731–871231          | 186         | 1.57          | 1.51     | 3.8    | 1.51     | 4.3    | 1.55     | 1.2    | 1.57      | −0.2   |
| <i>Value-Weighted:</i> |             |               |          |        |          |        |          |        |           |        |
| 260130–871231          | 744         | 1.33          | 1.27     | 4.5    | 1.26     | 5.5    | 1.22     | 8.4    | 1.19      | 11.1   |
| 260130–561231          | 372         | 1.57          | 1.51     | 4.5    | 1.51     | 4.3    | 1.44     | 9.5    | 1.38      | 14.5   |
| 570131–871231          | 372         | 1.28          | 1.22     | 4.4    | 1.18     | 7.9    | 1.21     | 5.6    | 1.24      | 2.7    |
| 260130–410630          | 186         | 1.57          | 1.52     | 3.2    | 1.55     | 1.0    | 1.49     | 5.5    | 1.42      | 10.6   |
| 410731–561231          | 186         | 1.26          | 1.18     | 6.4    | 1.11     | 12.9   | 1.07     | 17.1   | 1.08      | 16.1   |
| 570131–720630          | 186         | 1.05          | 0.96     | 9.3    | 0.92     | 14.7   | 0.95     | 10.9   | 1.01      | 4.7    |
| 720731–871231          | 186         | 1.51          | 1.48     | 1.6    | 1.45     | 4.0    | 1.47     | 2.4    | 1.49      | 1.1    |



**Figure 6.2.** Autocorrelograms of equally-weighted CRSP daily and monthly stock returns indexes and fractionally-differenced process with  $d = 1/4$ . The sample period for the daily index is July 1962 to December 1987, and is January 1926 to December 1987 for the monthly index.

indexes only the lowest order autocorrelation coefficients are statistically significant. For comparison, alongside each of the index's autocorrelogram is the autocorrelogram of the fractionally-differenced process (6.2.1) with  $d = .25$  and the variance of the disturbance chosen to yield a first-order autocorrelation of  $\frac{1}{3}$ . Although the general shapes of the fractionally-differenced autocorrelograms seem consistent with the data, closer inspection reveals that the index autocorrelations decay much more rapidly. Therefore, although short-term correlations are large enough to drive  $\tilde{Q}_n$  and  $Q_n$  apart, there is little evidence of long-range dependence in  $Q_n$  itself.

Additional results are available for weekly and annual stock returns data but since they are so similar to those reported here, we have omitted them to conserve space. Although the annual data spans 115 years (1872 to 1986), neither the classical nor the modified  $R/S$  statistics are statistically significant over this time span.

The evidence in Tables 6.3 and 6.4 shows that the null hypothesis of short-range dependence cannot be rejected by the data—there is little support for long-term memory in U.S. stock returns. With adjustments for autocorrelation at lags up to one calendar year, estimates of the modified rescaled range are consistent with the null hypothesis of weak dependence. This reinforces Kandel and Stambaugh's (1989) contention that the long-run predictability of stock returns uncovered by Fama and French (1988) and Poterba and Summers (1988) may not be "long-run" in the time series sense, but may be the result of more conventional models of short-range dependence.<sup>21</sup> Of course, since our inferences rely solely on asymptotic distribution theory, we must check our approximations before dismissing the possibility of long-range dependence altogether. The finite-sample size and power of the modified rescaled range test are considered in the next sections.

## 6.5 Size and Power

To explore the possibility that the inability to reject the null hypothesis of short-range dependence is merely a symptom of low power, and to check the quality of Section 6.3's asymptotic approximations for various sample sizes, we perform several illustrative Monte Carlo experiments. Section 6.5.1 reports the empirical size of the test statistic under two Gaussian null hypotheses: IID and AR(1) disturbances. Section 6.5.2 presents power results against the fractionally-differenced process (6.2.1) for  $d = \frac{1}{3}$  and  $-\frac{1}{3}$ .

### 6.5.1 The Size of the $R/S$ Test

Table 6.5a contains simulation results for the modified  $R/S$  statistic with sample sizes of 100, 250, 500, 750, and 1,000 under the null hypothesis of independently and identically distributed Gaussian errors. All simulations were performed on an IBM 4381 in double precision using the random generator G05DDF from the Numerical Algorithms Group Fortran Library Mark 12. For each sample size the statistic  $V_n(q)$  is computed with  $q = 0, 5, 10, 25, 50$ , and with  $q$  chosen by Andrews' (1991) data-dependent formula:

$$q = [k_n], \quad k_n \equiv \left(\frac{3n}{2}\right)^{\frac{1}{3}} \cdot \left(\frac{2\hat{\rho}}{1 - \hat{\rho}^2}\right)^{\frac{2}{3}}, \quad (6.5.1)$$

<sup>21</sup>Moreover, several papers have suggested that these long-run results may be spurious. See, for example, Kim, Nelson, and Startz (1991), Richardson (1993), and Richardson and Stock (1990).

where  $[k_n]$  denotes the greatest integer less than or equal to  $k_n$ , and  $\hat{\rho}$  is the estimated first-order autocorrelation coefficient of the data.<sup>22</sup> (Note that this is an optimal truncation lag only for an AR(1) data-generating process—a different expression obtains if, for example, the data-generating process were assumed to be an ARMA(1,1). See Andrews (1991) for further details.) In this case, the entry reported in the column labelled “ $q$ ” is the mean of the  $q$ ’s chosen, with the population standard deviation reported in parentheses below the mean. When  $q = 0$ ,  $V_n(q)$  is identical to Mandelbrot’s classical  $R/S$  statistic  $\tilde{V}_n$ .

The entries in the last three columns of Table 6.5a show that the classical  $R/S$  statistic tends to reject too frequently—even for sample sizes of 1,000 the empirical size of a 5 percent test based on  $\tilde{V}_n$  is 5.9 percent. The modified  $R/S$  statistic tends to be conservative for values of  $q$  that are not too large relative to the sample size. For example, with 100 observations and 5 lags the empirical size of the 5 percent test using  $V_n(q)$  is 2.1 percent. However, with 50 lags this test has a rejection rate of 31 percent! That the sampling properties worsen with the number of lags is not surprising—the imprecision with which the higher-order autocovariances are estimated can introduce considerable noise into the statistic (see, for example, Lo and MacKinlay, 1989b). But for 1,000 observations and 5 lags, the size of a 5 percent test based on  $V_n(q)$  is 5.1 percent. Andrews’ procedure yields intermediate results, with sizes in between those of the classical  $R/S$  statistic and the closest of the modified  $R/S$  statistics.

Table 6.5b reports the results of simulations under the null hypothesis of a Gaussian AR(1) with autoregressive coefficient 0.5 (recall that such a process is weakly dependent). The last three columns confirm the example of Section 6.3 and accord well with the results of Davies and Harte (1987): tests based on the classical  $R/S$  statistic have considerable power against an AR(1) null. In samples of only 100 observations the empirical size of the 5 percent test based on  $\tilde{V}_n$  is 38 percent and increases to 62 percent for sample sizes of 1,000. In contrast, the empirical sizes of tests based on  $V_n(q)$  are much closer to their nominal values since the geometrically declining autocorrelations are taken into account by the denominator  $\hat{\sigma}_n(q)$  of  $V_n(q)$ . When  $q$  is chosen via Andrews’ procedure, this yields conservative test sizes, ranging from 2.8 percent for a sample of 100, to 4.3 percent for a sample of 1,000.

<sup>22</sup>For this procedure, the Newey-West autocorrelation weights (6.3.7) are replaced by those suggested by Andrews (1991):

$$\omega_j = 1 - \left| \frac{j}{k_n} \right|.$$

**Table 6.5a.** Finite sample distribution of the modified R/S statistic under an IID null hypothesis. Each set of rows of a given sample size  $n$  corresponds to a separate and independent Monte Carlo experiment based on 10,000 replications. A lag  $q$  of 0 corresponds to Mandelbrot's classical R/S statistic, and a noninteger lag value indicates the mean lag (standard deviation given in parentheses) chosen via Andrews' (1991) data-dependent procedure assuming an AR(1) data-generating process. Standard errors for the empirical size may be computed using the usual normal approximation: they are  $9.95 \times 10^{-4}$ ,  $2.18 \times 10^{-3}$ , and  $3.00 \times 10^{-3}$  for the 1%, 5%, and 10% tests respectively.

| $n$ | $q$            | Min   | Max   | Mean  | S.D.  | Size<br>1%-Test | Size<br>5%-Test | Size<br>10%-Test |
|-----|----------------|-------|-------|-------|-------|-----------------|-----------------|------------------|
| 100 | 0              | 0.534 | 2.284 | 1.144 | 0.263 | 0.029           | 0.095           | 0.153            |
| 100 | 5              | 0.649 | 1.913 | 1.179 | 0.207 | 0.002           | 0.021           | 0.050            |
| 100 | 10             | 0.710 | 1.877 | 1.223 | 0.175 | 0.000           | 0.003           | 0.012            |
| 100 | 25             | 0.858 | 2.296 | 1.383 | 0.186 | 0.001           | 0.014           | 0.039            |
| 100 | 50             | 0.918 | 3.119 | 1.694 | 0.360 | 0.137           | 0.313           | 0.414            |
| 100 | 0.97<br>(0.83) | 0.557 | 2.164 | 1.150 | 0.247 | 0.019           | 0.070           | 0.127            |
| 250 | 0              | 0.496 | 2.527 | 1.183 | 0.270 | 0.021           | 0.075           | 0.133            |
| 250 | 5              | 0.580 | 2.283 | 1.196 | 0.243 | 0.008           | 0.041           | 0.089            |
| 250 | 10             | 0.654 | 2.048 | 1.211 | 0.221 | 0.003           | 0.021           | 0.054            |
| 250 | 25             | 0.757 | 1.905 | 1.264 | 0.176 | 0.000           | 0.001           | 0.006            |
| 250 | 50             | 0.877 | 2.206 | 1.372 | 0.169 | 0.000           | 0.005           | 0.020            |
| 250 | 0.97<br>(0.83) | 0.497 | 2.442 | 1.185 | 0.263 | 0.017           | 0.064           | 0.120            |
| 500 | 0              | 0.518 | 2.510 | 1.201 | 0.267 | 0.015           | 0.061           | 0.117            |
| 500 | 5              | 0.589 | 2.357 | 1.207 | 0.252 | 0.008           | 0.047           | 0.094            |
| 500 | 10             | 0.630 | 2.227 | 1.215 | 0.240 | 0.004           | 0.032           | 0.073            |
| 500 | 25             | 0.677 | 2.051 | 1.240 | 0.210 | 0.000           | 0.008           | 0.029            |
| 500 | 50             | 0.709 | 1.922 | 1.285 | 0.176 | 0.000           | 0.001           | 0.005            |
| 500 | 0.96<br>(0.82) | 0.549 | 2.510 | 1.202 | 0.263 | 0.014           | 0.057           | 0.112            |
| 750 | 0              | 0.558 | 2.699 | 1.207 | 0.270 | 0.014           | 0.061           | 0.120            |
| 750 | 5              | 0.597 | 2.711 | 1.212 | 0.260 | 0.009           | 0.049           | 0.101            |
| 750 | 10             | 0.615 | 2.553 | 1.217 | 0.251 | 0.006           | 0.039           | 0.087            |
| 750 | 25             | 0.677 | 2.279 | 1.235 | 0.228 | 0.001           | 0.017           | 0.052            |
| 750 | 50             | 0.758 | 1.971 | 1.266 | 0.198 | 0.000           | 0.002           | 0.015            |
| 750 | 0.96<br>(0.83) | 0.558 | 2.670 | 1.208 | 0.268 | 00.013          | 0.058           | 0.117            |

(continued)

Table 6.5a. (continued)

| $n$  | $q$            | Min   | Max   | Mean  | S.D.  | Size<br>1%-Test | Size<br>5%-Test | Size<br>10%-Test |
|------|----------------|-------|-------|-------|-------|-----------------|-----------------|------------------|
| 1000 | 0              | 0.542 | 2.577 | 1.211 | 0.270 | 0.014           | 0.059           | 0.113            |
| 1000 | 5              | 0.566 | 2.477 | 1.214 | 0.262 | 0.011           | 0.051           | 0.103            |
| 1000 | 10             | 0.570 | 2.405 | 1.218 | 0.256 | 0.008           | 0.045           | 0.089            |
| 1000 | 25             | 0.616 | 2.203 | 1.231 | 0.237 | 0.003           | 0.025           | 0.061            |
| 1000 | 50             | 0.716 | 2.036 | 1.253 | 0.211 | 0.000           | 0.007           | 0.029            |
| 1000 | 0.96<br>(0.81) | 0.549 | 2.546 | 1.212 | 0.268 | 0.012           | 0.056           | 0.111            |

### 6.5.2 Power Against Fractionally-Differenced Alternatives

Tables 6.6a and b report the power of the  $R/S$  tests against the Gaussian fractionally-differenced alternative:

$$(1 - L)^d \epsilon_t = \eta_t, \quad \eta_t \text{ IID } \mathcal{N}(0, \sigma_\eta^2), \quad (6.5.2)$$

with  $d = \frac{1}{3}$  and  $-\frac{1}{3}$ , and  $\sigma_\eta^2 = \Gamma^2(1-d)/\Gamma(1-2d)$  so as to yield a unit variance for  $\epsilon_t$ . For sample sizes of 100, tests based on  $V_n(q)$  have very little power, but when the sample size reaches 250 the power increases dramatically. According to Table 6.6a, the power of the 5 percent test with  $q = 5$  against the  $d = \frac{1}{3}$  alternative is 33.5 percent with 250 observations, 62.8 percent with 500 observations, and 84.6 percent with 1,000 observations. Although Andrews' automatic truncation lag procedure is generally less powerful, its power is still 63.0 percent for a sample size of 1,000. Also, the rejections are generally in the right tail of the distribution, as the entries in the "Max" column indicate. This is not surprising in light of Theorem 6.3.3, which shows that under this alternative the modified  $R/S$  statistic diverges in probability to infinity.

For a fixed sample size, the power of the  $V_n(q)$ -based test declines as the number of lags is increased. This is due to the denominator  $\hat{\sigma}_n(q)$ , which generally increases with  $q$  since there is positive dependence when  $d = \frac{1}{3}$ . The increase in the denominator decreases the mean and variance of the statistic, shifting the distribution towards the left and pulling probability mass from both tails, thereby reducing the frequency of draws in the right tail's critical region, where virtually all the power is coming from.

Against the  $d = -\frac{1}{3}$  alternative, Table 6.6b shows that the test seems to have somewhat higher power. However, in contrast to Table 6.6a the rejections are now coming from the left tail of the distribution, as Theorem 6.3.3 predicts. Although less powerful, tests based on Andrews' procedure

**Table 6.5b.** Finite sample distribution of the modified  $R/S$  statistic under an  $AR(1)$  null hypothesis with autoregressive coefficient 0.5. Each set of rows of a given sample size  $n$  corresponds to a separate and independent Monte Carlo experiment based on 10,000 replications. A lag  $q$  of 0 corresponds to Mandelbrot's classical  $R/S$  statistic, and a noninteger lag value indicates the mean lag (standard deviation given in parentheses) chosen via Andrews' (1991) data-dependent procedure, assuming an  $AR(1)$  data-generating process. Standard errors for the empirical size may be computed using the usual normal approximation; they are  $9.95 \times 10^{-4}$ ,  $2.18 \times 10^{-3}$ , and  $3.00 \times 10^{-3}$  for the 1%, 5%, and 10% tests respectively.

| $n$ | $q$             | Min   | Max   | Mean  | S.D.  | Size<br>1%-Test | Size<br>5%-Test | Size<br>10%-Test |
|-----|-----------------|-------|-------|-------|-------|-----------------|-----------------|------------------|
| 100 | 0               | 0.764 | 3.418 | 1.764 | 0.402 | 0.203           | 0.382           | 0.486            |
| 100 | 5               | 0.634 | 1.862 | 1.201 | 0.220 | 0.003           | 0.027           | 0.059            |
| 100 | 10              | 0.693 | 1.805 | 1.178 | 0.176 | 0.000           | 0.010           | 0.030            |
| 100 | 25              | 0.779 | 2.111 | 1.290 | 0.175 | 0.000           | 0.005           | 0.015            |
| 100 | 50              | 0.879 | 3.013 | 1.571 | 0.341 | 0.074           | 0.198           | 0.284            |
| 100 | 5.61<br>(1.25)  | 0.636 | 1.974 | 1.195 | 0.219 | 0.004           | 0.028           | 0.063            |
| 250 | 0               | 0.865 | 3.720 | 1.913 | 0.432 | 0.309           | 0.505           | 0.614            |
| 250 | 5               | 0.597 | 2.478 | 1.268 | 0.262 | 0.005           | 0.038           | 0.086            |
| 250 | 10              | 0.615 | 2.137 | 1.212 | 0.228 | 0.003           | 0.023           | 0.063            |
| 250 | 25              | 0.734 | 1.811 | 1.218 | 0.177 | 0.000           | 0.004           | 0.015            |
| 250 | 50              | 0.809 | 2.119 | 1.304 | 0.166 | 0.000           | 0.003           | 0.010            |
| 250 | 8.07<br>(1.07)  | 0.603 | 2.357 | 1.227 | 0.242 | 0.004           | 0.030           | 0.071            |
| 500 | 0               | 0.836 | 4.392 | 1.980 | 0.456 | 0.363           | 0.559           | 0.665            |
| 500 | 5               | 0.622 | 2.557 | 1.302 | 0.285 | 0.012           | 0.055           | 0.109            |
| 500 | 10              | 0.579 | 2.297 | 1.236 | 0.256 | 0.007           | 0.039           | 0.085            |
| 500 | 25              | 0.627 | 1.980 | 1.214 | 0.215 | 0.001           | 0.015           | 0.041            |
| 500 | 50              | 0.734 | 1.894 | 1.243 | 0.178 | 0.000           | 0.002           | 0.009            |
| 500 | 10.40<br>(0.99) | 0.577 | 2.353 | 1.236 | 0.256 | 0.007           | 0.039           | 0.085            |
| 750 | 0               | 0.839 | 4.211 | 2.017 | 0.459 | 0.389           | 0.592           | 0.696            |
| 750 | 5               | 0.567 | 2.637 | 1.323 | 0.291 | 0.011           | 0.062           | 0.118            |
| 750 | 10              | 0.557 | 2.429 | 1.253 | 0.265 | 0.007           | 0.043           | 0.091            |
| 750 | 25              | 0.614 | 2.114 | 1.222 | 0.232 | 0.003           | 0.022           | 0.058            |
| 750 | 50              | 0.702 | 1.891 | 1.235 | 0.200 | 0.000           | 0.005           | 0.022            |
| 750 | 12.03<br>(0.93) | 0.556 | 2.324 | 1.244 | 0.260 | 0.007           | 0.041           | 0.088            |

(continued)

Table 6.5b. (continued)

| $n$  | $q$             | Min   | Max   | Mean  | S.D.  | Size<br>1%-Test | Size<br>5%-Test | Size<br>10%-Test |
|------|-----------------|-------|-------|-------|-------|-----------------|-----------------|------------------|
| 1000 | 0               | 0.926 | 4.327 | 2.045 | 0.465 | 0.414           | 0.617           | 0.716            |
| 1000 | 5               | 0.625 | 2.768 | 1.340 | 0.296 | 0.014           | 0.065           | 0.125            |
| 1000 | 10              | 0.592 | 2.622 | 1.268 | 0.272 | 0.009           | 0.047           | 0.096            |
| 1000 | 25              | 0.608 | 2.350 | 1.231 | 0.244 | 0.004           | 0.030           | 0.072            |
| 1000 | 50              | 0.636 | 1.997 | 1.236 | 0.217 | 0.001           | 0.011           | 0.038            |
| 1000 | 13.30<br>(0.89) | 0.590 | 2.548 | 1.252 | 0.265 | 0.008           | 0.043           | 0.090            |

still exhibit reasonable power, ranging from 33.1 percent in samples of 100 observations to 94.5 percent in samples of 1,000.

For the larger sample sizes the power again declines as the number of lags increases, due to the denominator  $\hat{\sigma}_n(q)$ , which declines as  $q$  increases because the population autocorrelations are all negative when  $d = -\frac{1}{3}$ . The resulting increase in the mean of  $V_n(q)$ 's sampling distribution overwhelms the increase in its variability, leading to a lower rejection rate from the left tail.

Table 6.6a and b shows that the modified  $R/S$  statistic has reasonable power against at least two specific models of long-term memory. However, these simulations are merely illustrative—a more conclusive study would include further simulations with several other values for  $d$ , and perhaps with short-range dependence as well.<sup>23</sup> Moreover, since our empirical work has

<sup>23</sup>The very fact that the modified  $R/S$  statistic yields few rejections under the null simulations of Section 6.5.1 shows that the test may have low power against *some* long-range dependent alternatives, since the pseudo-random number generator used in those simulations is, after all, a long-range dependent process. A more striking example is the “tent” map, a particularly simple nonlinear deterministic map (it has a correlation dimension of 1) which yields sequences that are virtually uncorrelated but long-range dependent. In particular, the tent map is given by the following recursion:

$$X_t = \begin{cases} 2X_{t-1} & \text{if } X_{t-1} < \frac{1}{2} \\ 2(1 - X_{t-1}) & \text{if } X_{t-1} \geq \frac{1}{2}, \end{cases} \quad t = 1, \dots, T, \quad X_0 \in (0, 1).$$

As an illustration, we performed two Monte Carlo experiments using the tent map to generate samples of 500 and 1,000 observations (each with 10,000 replications) with an independent uniform (0,1) starting value for each replication. Neither the Mandelbrot rescaled range, nor its modification with fixed or automatic truncation lags have any power against the tent map. In fact, the finite sample distributions are quite close to the null distribution. Of course, one could argue that if the dynamics and the initial condition were unknown, then even if a deterministic system *were* generating the data, the resulting time series would be short-range dependent “for all practical purposes” and *should* be part of our null. We are grateful to Lars Hansen for suggesting this analysis.

**Table 6.6a.** Power of the modified R/S statistic under a Gaussian fractionally differenced alternative with differencing parameter  $d = 1/3$ . The variance of the process has been normalized to unity. Each set of rows of a given sample size  $n$  corresponds to a separate and independent Monte Carlo experiment based on 10,000 replications. A lag  $q$  of 0 corresponds to Mandelbrot's classical R/S statistic, and a noninteger lag value indicates the mean lag (standard deviation given in parentheses) chosen via Andrews' (1991) data-dependent procedure assuming an AR(1) data-generating process.

| $n$  | $q$             | Min   | Max   | Mean  | S.D.  | Power<br>1%-Test | Power<br>5%-Test | Power<br>10%-Test |
|------|-----------------|-------|-------|-------|-------|------------------|------------------|-------------------|
| 100  | 0               | 0.729 | 4.047 | 2.025 | 0.513 | 0.429            | 0.600            | 0.680             |
| 100  | 5               | 0.635 | 2.089 | 1.361 | 0.242 | 0.001            | 0.014            | 0.065             |
| 100  | 10              | 0.686 | 1.746 | 1.237 | 0.171 | 0.000            | 0.005            | 0.015             |
| 100  | 25              | 0.723 | 2.148 | 1.208 | 0.156 | 0.000            | 0.002            | 0.008             |
| 100  | 50              | 0.823 | 2.803 | 1.399 | 0.328 | 0.035            | 0.096            | 0.154             |
| 100  | 4.27<br>(1.43)  | 0.650 | 2.330 | 1.411 | 0.257 | 0.003            | 0.040            | 0.104             |
| 250  | 0               | 0.938 | 5.563 | 2.678 | 0.709 | 0.774            | 0.878            | 0.918             |
| 250  | 5               | 0.713 | 2.924 | 1.699 | 0.364 | 0.153            | 0.335            | 0.442             |
| 250  | 10              | 0.705 | 2.304 | 1.475 | 0.277 | 0.006            | 0.091            | 0.186             |
| 250  | 25              | 0.681 | 1.852 | 1.264 | 0.175 | 0.000            | 0.003            | 0.012             |
| 250  | 50              | 0.756 | 1.971 | 1.208 | 0.140 | 0.000            | 0.002            | 0.007             |
| 250  | 6.63<br>(1.36)  | 0.711 | 2.596 | 1.619 | 0.317 | 0.067            | 0.240            | 0.360             |
| 500  | 0               | 1.061 | 7.243 | 3.336 | 0.929 | 0.924            | 0.967            | 0.980             |
| 500  | 5               | 0.731 | 3.726 | 2.055 | 0.491 | 0.450            | 0.628            | 0.709             |
| 500  | 10              | 0.692 | 2.944 | 1.750 | 0.384 | 0.197            | 0.385            | 0.494             |
| 500  | 25              | 0.623 | 2.164 | 1.429 | 0.258 | 0.001            | 0.045            | 0.123             |
| 500  | 50              | 0.687 | 1.763 | 1.271 | 0.178 | 0.000            | 0.003            | 0.009             |
| 500  | 8.96<br>(1.37)  | 0.709 | 3.201 | 1.809 | 0.384 | 0.242            | 0.447            | 0.557             |
| 750  | 0               | 1.228 | 8.059 | 3.799 | 1.052 | 0.972            | 0.990            | 0.995             |
| 750  | 5               | 0.838 | 4.280 | 2.313 | 0.565 | 0.620            | 0.766            | 0.830             |
| 750  | 10              | 0.769 | 3.421 | 1.955 | 0.447 | 0.370            | 0.557            | 0.655             |
| 750  | 25              | 0.734 | 2.478 | 1.569 | 0.310 | 0.042            | 0.195            | 0.304             |
| 750  | 50              | 0.722 | 1.925 | 1.359 | 0.223 | 0.000            | 0.005            | 0.036             |
| 750  | 10.58<br>(1.33) | 0.798 | 3.324 | 1.942 | 0.421 | 0.363            | 0.559            | 0.657             |
| 1000 | 0               | 1.398 | 8.615 | 4.174 | 1.174 | 0.985            | 0.996            | 0.998             |
| 1000 | 5               | 0.898 | 4.672 | 2.521 | 0.635 | 0.720            | 0.846            | 0.892             |
| 1000 | 10              | 0.779 | 3.766 | 2.121 | 0.504 | 0.494            | 0.669            | 0.747             |

(continued)

**Table 6.6a.** (continued)

| $n$  | $q$             | Min   | Max   | Mean  | S.D.  | Power<br>1%-Test | Power<br>5%-Test | Power<br>10%-Test |
|------|-----------------|-------|-------|-------|-------|------------------|------------------|-------------------|
| 1000 | 25              | 0.641 | 2.734 | 1.686 | 0.354 | 0.135            | 0.322            | 0.431             |
| 1000 | 50              | 0.628 | 2.118 | 1.441 | 0.259 | 0.001            | 0.052            | 0.138             |
| 1000 | 11.87<br>(1.31) | 0.766 | 3.613 | 2.044 | 0.454 | 0.446            | 0.630            | 0.718             |

**Table 6.6b.** Power of the modified R/S statistic under a Gaussian fractionally differenced alternative with differencing parameter  $d = -1/3$ . The variance of the process has been normalized to unity. Each set of rows of a given sample size  $n$  corresponds to a separate and independent Monte Carlo experiment based on 10,000 replications. A lag  $q$  of 0 corresponds to Mandelbrot's classical R/S statistic, and a noninteger lag value indicates the mean lag (standard deviation given in parentheses) chosen via Andrews' (1991) data-dependent procedure assuming an AR(1) data-generating process.

| $n$ | $q$            | Min   | Max   | Mean  | S.D.  | Power<br>1%-Test | Power<br>5%-Test | Power<br>10%-Test |
|-----|----------------|-------|-------|-------|-------|------------------|------------------|-------------------|
| 100 | 0              | 0.367 | 1.239 | 0.678 | 0.120 | 0.670            | 0.858            | 0.923             |
| 100 | 5              | 0.637 | 1.710 | 1.027 | 0.153 | 0.006            | 0.054            | 0.134             |
| 100 | 10             | 0.762 | 2.030 | 1.217 | 0.161 | 0.000            | 0.001            | 0.005             |
| 100 | 25             | 0.953 | 2.638 | 1.587 | 0.207 | 0.014            | 0.095            | 0.211             |
| 100 | 50             | 1.052 | 3.478 | 2.033 | 0.354 | 0.425            | 0.679            | 0.785             |
| 100 | 2.94<br>(0.99) | 0.478 | 1.621 | 0.889 | 0.155 | 0.131            | 0.331            | 0.466             |
| 250 | 0              | 0.303 | 1.014 | 0.561 | 0.089 | 0.951            | 0.991            | 0.997             |
| 250 | 5              | 0.549 | 1.479 | 0.851 | 0.128 | 0.146            | 0.409            | 0.571             |
| 250 | 10             | 0.632 | 1.752 | 1.005 | 0.143 | 0.007            | 0.065            | 0.152             |
| 250 | 25             | 0.833 | 1.936 | 1.292 | 0.157 | 0.000            | 0.001            | 0.004             |
| 250 | 50             | 0.977 | 2.357 | 1.594 | 0.186 | 0.007            | 0.078            | 0.198             |
| 250 | 4.20<br>(0.86) | 0.448 | 1.437 | 0.796 | 0.129 | 0.301            | 0.578            | 0.716             |
| 500 | 0              | 0.292 | 0.819 | 0.479 | 0.071 | 0.997            | 1.000            | 1.000             |
| 500 | 5              | 0.458 | 1.244 | 0.728 | 0.105 | 0.517            | 0.794            | 0.888             |
| 500 | 10             | 0.555 | 1.489 | 0.861 | 0.121 | 0.111            | 0.366            | 0.543             |
| 500 | 25             | 0.706 | 1.735 | 1.105 | 0.143 | 0.000            | 0.004            | 0.022             |
| 500 | 50             | 0.881 | 2.089 | 1.356 | 0.157 | 0.000            | 0.002            | 0.011             |
| 500 | 5.45<br>(0.77) | 0.443 | 1.318 | 0.725 | 0.108 | 0.529            | 0.793            | 0.887             |

(continued)

Table 6.6b. (continued)

| <i>n</i> | <i>q</i>       | Min   | Max   | Mean  | S.D.  | Power<br>1%-Test | Power<br>5%-Test | Power<br>10%-Test |
|----------|----------------|-------|-------|-------|-------|------------------|------------------|-------------------|
| 750      | 0              | 0.276 | 0.700 | 0.433 | 0.063 | 1.000            | 1.000            | 1.000             |
| 750      | 5              | 0.422 | 1.070 | 0.659 | 0.094 | 0.764            | 0.932            | 0.973             |
| 750      | 10             | 0.499 | 1.262 | 0.779 | 0.109 | 0.325            | 0.641            | 0.789             |
| 750      | 25             | 0.689 | 1.570 | 1.001 | 0.132 | 0.003            | 0.049            | 0.138             |
| 750      | 50             | 0.837 | 1.802 | 1.227 | 0.148 | 0.000            | 0.000            | 0.002             |
| 750      | 6.34<br>(0.73) | 0.424 | 1.133 | 0.682 | 0.099 | 0.679            | 0.892            | 0.951             |
| 1000     | 0              | 0.257 | 0.775 | 0.403 | 0.057 | 1.000            | 1.000            | 1.000             |
| 1000     | 5              | 0.401 | 1.149 | 0.613 | 0.085 | 0.895            | 0.978            | 0.993             |
| 1000     | 10             | 0.487 | 1.376 | 0.725 | 0.099 | 0.525            | 0.809            | 0.907             |
| 1000     | 25             | 0.633 | 1.596 | 0.930 | 0.121 | 0.020            | 0.154            | 0.306             |
| 1000     | 50             | 0.778 | 1.820 | 1.139 | 0.139 | 0.000            | 0.000            | 0.006             |
| 1000     | 7.01<br>(0.70) | 0.412 | 1.235 | 0.651 | 0.092 | 0.789            | 0.945            | 0.978             |

employed data sampled at different frequencies (implying different values of  $d$  for different sample sizes), the trade-off between the time span of the data and the frequency of observation for the test's power may be an important issue. Nevertheless, the simulation results suggest that short-range dependence may be the more significant feature of recent stock market returns.

## 6.6 Conclusion

Using a simple modification of the Hurst-Mandelbrot rescaled range that accounts for short-term dependence, and contrary to previous studies, we find little evidence of long-term memory in historical U.S. stock market returns. If the source of serial correlation is lagged adjustment to new information, the absence of strong dependence in stock returns should not be surprising from an economic standpoint, given the frequency with which financial asset markets clear. Surely financial security prices must be immune to persistent informational asymmetries, especially over longer time spans. Perhaps the fluctuations of aggregate economic output are more likely to display such long-run tendencies, as Kondratiev and Kuznets have suggested, and this long-memory in output may eventually manifest itself in the return to equity. But if some form of long-range dependence is indeed present in stock returns, it will not be easily detected by any of

Table 6.6b. (continued)

| <i>n</i> | <i>q</i>       | Min   | Max   | Mean  | S.D.  | Power<br>1%-Test | Power<br>5%-Test | Power<br>10%-Test |
|----------|----------------|-------|-------|-------|-------|------------------|------------------|-------------------|
| 750      | 0              | 0.276 | 0.700 | 0.433 | 0.063 | 1.000            | 1.000            | 1.000             |
| 750      | 5              | 0.422 | 1.070 | 0.659 | 0.094 | 0.764            | 0.932            | 0.973             |
| 750      | 10             | 0.499 | 1.262 | 0.779 | 0.109 | 0.325            | 0.641            | 0.789             |
| 750      | 25             | 0.689 | 1.570 | 1.001 | 0.132 | 0.003            | 0.049            | 0.138             |
| 750      | 50             | 0.837 | 1.802 | 1.227 | 0.148 | 0.000            | 0.000            | 0.002             |
| 750      | 6.34<br>(0.73) | 0.424 | 1.133 | 0.682 | 0.099 | 0.679            | 0.892            | 0.951             |
| 1000     | 0              | 0.257 | 0.775 | 0.403 | 0.057 | 1.000            | 1.000            | 1.000             |
| 1000     | 5              | 0.401 | 1.149 | 0.613 | 0.085 | 0.895            | 0.978            | 0.993             |
| 1000     | 10             | 0.487 | 1.376 | 0.725 | 0.099 | 0.525            | 0.809            | 0.907             |
| 1000     | 25             | 0.633 | 1.596 | 0.930 | 0.121 | 0.020            | 0.154            | 0.306             |
| 1000     | 50             | 0.778 | 1.820 | 1.139 | 0.139 | 0.000            | 0.000            | 0.006             |
| 1000     | 7.01<br>(0.70) | 0.412 | 1.235 | 0.651 | 0.092 | 0.789            | 0.945            | 0.978             |

employed data sampled at different frequencies (implying different values of  $d$  for different sample sizes), the trade-off between the time span of the data and the frequency of observation for the test's power may be an important issue. Nevertheless, the simulation results suggest that short-range dependence may be the more significant feature of recent stock market returns.

## 6.6 Conclusion

Using a simple modification of the Hurst-Mandelbrot rescaled range that accounts for short-term dependence, and contrary to previous studies, we find little evidence of long-term memory in historical U.S. stock market returns. If the source of serial correlation is lagged adjustment to new information, the absence of strong dependence in stock returns should not be surprising from an economic standpoint, given the frequency with which financial asset markets clear. Surely financial security prices must be immune to persistent informational asymmetries, especially over longer time spans. Perhaps the fluctuations of aggregate economic output are more likely to display such long-run tendencies, as Kondratiev and Kuznets have suggested, and this long-memory in output may eventually manifest itself in the return to equity. But if some form of long-range dependence is indeed present in stock returns, it will not be easily detected by any of

our current statistical tools, especially in view of the optimality of the  $R/S$  statistic in the Mandelbrot and Wallis (1969b) sense. Direct estimation of particular parametric models may provide more positive evidence of long-term memory and is currently being pursued by several investigators.<sup>24</sup>

<sup>24</sup>See, for example, Boes et al. (1989), Diebold and Rudebusch (1989), Fox and Taqqu (1986), Geweke and Porter-Hudak (1983), Porter-Hudak (1990), Sowell (1989, 1990), and Yajima (1985, 1988).

## Appendix A6

### Proof of Theorems

Proofs of the theorems rely on the following three lemmas:

**Lemma A.1 (Herrndorf (1984)).** *If  $\{\epsilon_t\}$  satisfies assumptions (A1)–(A4) then as  $n$  increases without bound,  $W_n(\tau) \Rightarrow W(\tau)$ .*

**Lemma A.2 (Extended Continuous Mapping Theorem).**<sup>25</sup> *Let  $h_n$  and  $h$  be measurable mappings from  $\mathcal{D}[0, 1]$  to itself and denote by  $E$  the set of  $x \in \mathcal{D}[0, 1]$  such that  $h_n(x_n) \rightarrow h(x)$  fails to hold for some sequence  $x_n$  converging to  $x$ . If  $W_n(\tau) \Rightarrow W(\tau)$  and  $E$  is of Wiener-measure zero, i.e.,  $P(W \in E) = 0$ , then  $h_n(W_n) \Rightarrow h(W)$ .*

**Lemma A.3.** *Let  $R_n \Rightarrow R$  where both  $R_n$  and  $R$  have nonnegative support, and let  $P(R = 0) = P(R = \infty) = 0$ . If  $a_n \xrightarrow{p} \infty$ , then  $a_n R_n \xrightarrow{p} \infty$ . If  $a_n \xrightarrow{p} 0$ , then  $a_n R_n \xrightarrow{p} 0$ .*

#### Proof of Theorem 6.3.1

Let  $S_n = \sum_{j=1}^n \epsilon_j$  and define the following function  $Y_n(\tau)$  on  $\mathcal{D}[0, 1]$ :

$$Y_n(\tau) = \frac{1}{\sigma \sqrt{n}} S_{[n\tau]}, \quad \tau \in [0, 1], \quad (\text{A6.1})$$

where  $[n\tau]$  denotes the greatest integer less than or equal to  $n\tau$ , and  $\sigma$  is defined in condition (A3) of the null hypothesis. By convention, set  $Y_n(0) \equiv 0$ . Under conditions (A1), (A2'), (A3), and (A4) Herrndorf (1984)

<sup>25</sup>See Billingsley (1968) for a proof.

has shown that  $Y_n(\tau) \Rightarrow W(\tau)$ . But consider:

$$\begin{aligned} \text{Max}_{1 \leq k \leq n} \frac{1}{\hat{\sigma}_n(q)\sqrt{n}} \sum_{j=1}^k (X_j - \bar{X}_n) \\ = \text{Max}_{1 \leq k \leq n} \frac{1}{\hat{\sigma}_n(q)\sqrt{n}} \left( S_k - \frac{k}{n} S_n \right) \end{aligned} \quad (\text{A6.2a})$$

$$= \text{Max}_{0 \leq \tau \leq 1} Z_n(\tau) \quad (\text{A6.2b})$$

where

$$Z_n(\tau) \equiv Y_n(\tau) - \frac{[n\tau]}{n} Y_n(1). \quad (\text{A6.2c})$$

Since the sequence of functions  $h_n$  that map  $Y_n(\tau)$  to  $Z_n(\tau)$  satisfies the conditions of Lemma A.2, where the limiting mapping  $h$  takes  $Y_n(\tau)$  to  $Y_n(\tau) - \tau Y_n(1)$ , it may be concluded that

$$h_n(Y_n(\tau)) = Z_n(\tau) \Rightarrow h(W(\tau)) = W(\tau) - \tau W(1) = W^0(\tau). \quad (\text{A6.3})$$

If the estimator  $\hat{\sigma}_n(q)$  is substituted in place of  $\sigma$  in the construction of  $Z_n(\tau)$ , then under conditions (A2') and (A5), Theorem 4.2 of Phillips (1987) shows that (A3) still obtains. The rest of the theorem follows directly from repeated application of Lemma A.2. **Q.E.D.**

### Proof of Theorem 6.3.2

See Davydov (1970) and Taqqu (1975).

### Proof of Theorem 6.3.3

Parts (6.3.3)–(6.3.3) follow directly from Theorem 6.3.2 and Lemma A.2, and part (6.3.3) follows immediately from Lemma A.3. Therefore, we need only prove (6.3.3). Let  $H \in (\frac{1}{2}, 1)$  so that  $\gamma(k) \sim k^{2H-2} L(k)$ . This implies that

$$\text{Var}[S_n] \sim n^{2H} L(n). \quad (\text{A6.4})$$

Therefore, to show that  $a_n \xrightarrow{p} \infty$ , it suffices to show that

$$\frac{\hat{\sigma}^2(q)}{n^{2H-1} L(n)} \xrightarrow{p} 0. \quad (\text{A6.5})$$

Consider the population counterpart to (A6.5):

$$\frac{\sigma^2(q)}{n^{2H-1} L(n)} = \frac{1}{n^{2H-1} L(n)} \left( \sigma_\epsilon^2 + 2 \sum_{j=1}^q \omega_j \gamma_j \right) \quad (\text{A6.6})$$

where  $\omega_j = 1 - j/(q+1)$ . Since by assumption  $\gamma_j \sim j^{2H-2}L(j)$ , there exists some integer  $q_a$  and  $M > 0$  such that for  $j > q_a$ ,  $\gamma_j < Mj^{2H-2}L(j)$ . Now it is well known that a slowly-varying function satisfies the inequality  $j^{-\epsilon} < L(j) < j^\epsilon$  for any  $\epsilon > 0$  and  $j > q_b$ , for some  $q_b(\epsilon)$ . Choose  $\epsilon < 2 - 2H$ , and observe that

$$\gamma_j < Mj^{2H-2}j^\epsilon, \quad j > q_0 \equiv \max(q_a, q_b) \quad (\text{A6.7})$$

which implies:

$$2 \sum_{j=1}^q \omega_j \gamma_j < 2 \sum_{j=1}^{q_0} \omega_j \gamma_j + 2M \sum_{j=q_0+1}^q \omega_j j^{2H-2+\epsilon} \quad (\text{A6.8})$$

where, without loss of generality, we have assumed that  $q > q_0$ . As  $q$  increases without bound, the first sum of the right-side of (A6.8) remains finite, and the second sum may be bounded by observing that its summands are positive and decreasing, hence (see, for example, Buck, 1978, Chapter 5.5):

$$2M \sum_{j=q_0+1}^q \omega_j j^{2H-2+\epsilon} \leq 2M \int_{q_0}^q \left(1 - \frac{x}{q+1}\right) x^{2H-2+\epsilon} dx \quad (\text{A6.9a})$$

$$\sim O(q^{2H-1+\epsilon}) \quad (\text{A6.9b})$$

where the asymptotic equivalence follows by direct integration. If  $q \sim O(n^\delta)$  where  $\delta \in (0, 1)$ , a weaker condition than required by our null hypothesis, then the ratio  $\sigma^2(q)/(n^{2H-1}L(n))$  is at most of order  $O(n^{(2H-1+\epsilon)(\delta-1)})$ , which converges to zero. If we can now show that (A6.6) and its sample counterpart are equal in probability, then we are done. This is accomplished by the following sequence of inequalities:

$$\mathbb{E} \left| \frac{\hat{\sigma}^2(q)}{n^{2H-1}} - \frac{\sigma^2(q)}{n^{2H-1}} \right| = \frac{1}{n^{2H-1}} \mathbb{E} \left| (\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2) + 2 \sum_{j=1}^q \omega_j (\hat{\gamma}_j - \gamma_j) \right| \quad (\text{A6.10a})$$

$$\begin{aligned} &\leq \frac{\mathbb{E}|\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|}{n^{2H-1}} \\ &\quad + \frac{2}{n^{2H-1}} \sum_{j=1}^q \omega_j \mathbb{E}|\hat{\gamma}_j - \gamma_j| \end{aligned} \quad (\text{A6.10b})$$

$$\begin{aligned}
&\leq \frac{\mathbb{E}|\hat{\sigma}_\epsilon^2 - \sigma_\epsilon^2|}{n^{2H-1}} \\
&\quad + \frac{2}{n^{2H-1}} \sum_{j=1}^q \omega_j \sqrt{\mathbb{E}(\hat{\gamma}_j - \gamma_j)^2}. \tag{A6.10c}
\end{aligned}$$

But since Hosking (1984, Theorem 2) provides rates of convergence for sample auto-covariances of stationary Gaussian processes satisfying (6.3.14), an integral evaluation similar to that in (A6.9a) shows that the sum in (A6.10c) vanishes asymptotically when  $q \sim o(n)$ . This completes the proof. Since the proof for  $H \in (0, \frac{1}{2})$  is similar, it is left to the reader. **Q.E.D.**