Chapter 2

Literature Review

As Hsieh (1991) notes,

> After the stock market crash of October 19, 1987, interest in nonlinear dynamics, especially deterministic chaotic dynamics, has increased in both the financial press and the academic literature. This has come about because the frequency of large moves in stock markets is greater than would be expected under a normal distribution. There are a number of possible explanations. A popular one is that the stock market is governed by chaotic dynamics.

However, interest in a topic related to chaos and nonlinear dynamics has been around since the early 1960's. This is the idea that the sequence of stock returns is a "fractal" time series.

2.1. "Fractal" Dynamics

The statistical paradigm against which the sequence of increments from a chaotic dynamical process is typically contrasted is that of a white noise process. "White noise" typically refers to a sequence whose increments are independently and identically distributed with zero mean and finite variance. Brownian motion, a well-known paradigm in finance, is a white noise process for which the independent increments are identically normally distributed. A fractal time series is one which is statistically "self-similar" (apart from scale) regardless of the time frame over which the increments of the series are observed. For example, a sequence of daily observations would exhibit similar statistical characteristics to sequences of weekly, monthly, or yearly observations, and the scale of the observations would be a direct function of the length of time involved. Brownian motion, as a white noise process which exhibits these time series properties, is as Schroeder (1991) notes, "the paradigm of random fractals."

However, during the 1960's Benoit Mandelbrot, "the father of fractal geometry", believed that securities returns did in fact follow a fractal time series but that Brownian motion was not an adequate statistical description of the true stochastic process generating securities returns. In order to resolve this inadequacy, Mandelbrot worked in two perpendicular directions to expand the class of fractal time series. One direction involved relaxing the assumption of finite variance, which introduces what Mandelbrot termed the "Noah Effect." The other direction entailed relaxing the independence assumption, thereby allowing for a "Joseph Effect."

The Joseph Effect (see Mandelbrot (1972)) is named after the biblical story in which Joseph prophesied that the residents of Egypt would face seven years of feast followed by seven year of famine. This effect denotes the property of certain time series to exhibit persistent behavior (such as years of flooding followed by years of drought along the Nile River basin) more frequently than would be expected if the series were completely random, but without exhibiting any significant short-term (Markovian) dependence. To describe such processes, Mandelbrot broadened the idea of Brownian motion into the class of stochastic processes called "fractional"

Brownian motion (fBm), or, using the noise analogy, "flicker noises." Fractional Brownian motions exhibit complex, though linear, long-term dependencies and are characterized by a parameter called the Hurst exponent (H), which denotes the level of long-range dependence in the data and generally ranges from 0 to 1. If $H=0.5$, then the fractional Brownian motion will have no long-term persistence, and the result is standard Brownian motion, or white noise. Also, the power spectrum (the Fourier transform of the autocovariance function) of the increments of fractional Brownian motion will be proportional to f^{β}, where $\beta=2H-1$, so that Brownian motion or white noise will have a flat spectrum. If 0<H<0.5, then the series will exhibit anti-persistence, as evidenced by a greater number of reversals and fewer and shorter trends than in a white noise series. Visually, a graph of such a series would appear more jagged than a random walk. On the other hand, if $0.5<H<1$, then the series will exhibit persistence, with fewer reversals and longer trends than the increments of Brownian motion. In this case, the graph would appear smoother than that of a random walk. Also the power spectrum for such a series would be proportional to f^{β}, where $\beta>0$, so that the series would be subject to long-range dependence.

Incidentally, if $\beta=0$ (i.e. the power spectrum is flat), then the series is called "white" noise. If $\beta=1$, then the series is called "pink" noise. If a sequence is generated from a white noise series, with the next element in the sequence being equal to the sum of all the previous white noise increments, then the sequence is called "brown" noise and exhibits a $\beta=2$, which is indicative of its long-range persistence. For example, if the series of log returns for a stock is a white noise process in which each element in the series is independent, then the series of log stock values will be brown noise, for which the log of the current stock price will be dependent on all the previous log returns in the history of the stock. Finally, if $\beta=3$, then the series is called "black" noise and is heavily dependent upon its distant past. As Bak and Chen (1991) note in their paper on "self-organized criticality," black noise processes could govern such natural and unnatural catastrophes as floods, droughts, and stock market crashes. "Moreover," as Schroeder (1991) notes, "because of their black spectra, such disasters would come in clusters."

In order to measure what sort of noise a time series most closely resembles, Mandelbrot developed a statistical technique called "rescaled range" analysis, which yields a measure of the Hurst exponent. This involves comparing a linear measure of the spread of the time series (a variation of its sample range) to a quadratic measure (its sample standard deviation). Using this method, Greene and Fielitz (1977) found considerable evidence of temporal dependence in daily stock returns for the period December 23, 1963 to November 29, 1968, after accounting for short-term linear dependencies (autocorrelation) within the data. In a more popular treatment, Peters (1991) estimates the Hurst exponent to be 0.778 for monthly returns on the S&P 500 from January 1950 to July 1988. For a sample of individual stocks, Peters found Hurst exponents ranging from 0.75 for Apple Computer down to 0.54 for Consolidated Edison. All of these values are greater than 0.5, indicating a greater persistence among stock returns than would be expected if stock prices followed a geometric Brownian motion process.

The Noah Effect (recalling the Biblical account of the great deluge), on the other hand, refers to the tendency of various time series with presumably independent increments, especially speculative time series, to exhibit abrupt and discontinuous changes. Of the two effects enumerated by Mandelbrot, his examination of and hypothesized explanation for the Noah Effect have elicited the greater attention in the finance literature. The existence of abrupt discontinuities in financial time series, combined with the empirical observation of both sample leptokurtosis and unstable sample variances, led Mandelbrot to develop the "stable Paretian hypothesis" (see Mandelbrot (1963) and Mandelbrot (1983)). Under this hypothesis, stock returns follow a stable Paretian distribution with a characteristic exponent (α) whose value is between one and two. A stable Paretian distribution is a member of the class of distributions that are "invariant" under addition. An example of such a distribution is the normal, or Gaussian, distribution, and its invariance property is illustrated by the fact that the sum of two normal random variables is itself a normal random variable. However, other than for the normal random variables, closed-form solutions for the density functions of stable Paretian r.v.'s generally do not exist, but the natural log of the characterisic function for such a variable is given by:

$$\ln \phi_X(t) = i\delta t - \gamma |t|^{\alpha} \left[1 + i\beta \left(\frac{t}{|t|} \right) \tan\left(\frac{\alpha\pi}{2} \right) \right],$$

where α is the characteristic exponent or kurtosis parameter for the population, β is the skewness parameter, γ is the scale parameter, and δ is the location parameter. If $\alpha=2$ and $\beta=0$, then the variable is normally distributed with mean δ and variance 2γ. In general, only in the normal case are all the moments of the distribution finite. If $1<\alpha<2$, then the tails of the distribution taper off too slowly for the variance to be finite, and only the mean of the distribution exists. If $\alpha\leq 1$, then the tails taper off so slowly that not even the mean exists, in which case all of the moments of the distribution are nonexistent. The typical example for such a distribution is the Cauchy distribution ($\alpha=1$, $\beta=0$), which is equivalent to a Student's t distribution with one degree of freedom.

Using the log price differences of daily cotton prices over various time frames as an example of a speculative time series, Mandelbrot (1963) uses basically visual techniques to make a number of interesting observations. First, "the empirical distributions of price changes are usually too 'peaked' to be relative to samples from Gaussian populations." Furthermore, while "the histograms of price changes are indeed unimodal and their central 'bells' remind one of the 'Gaussian ogive', ... there are typically so many 'outliers' that ogives fitted to the mean square of price changes are much lower and flatter than the distribution of the data themselves. The tails of the distributions of price changes are in fact so extraordinarily long that the sample second moments typically vary in an erratic fashion." These observations are consistent with the 'stable Paretian hypothesis' Mandelbrot proposes. Via the use of graphs displaying the empirical probabilities of observations in the tail areas of the distributions, Mandelbrot obtains a value for α of 1.7, while the skewness and location parameters are estimated to be zero. Comparing the actual data to the probabilities obtained from a stable Patetian distribution with these parameter values, Mandelbrot finds that they produce a fairly good fit, although the actual data do seem to exhibit a slight amount of negative (left) skewness. However, the data still exhibit one major anomaly which the stable Paretian hypothesis cannot directly explain—"large changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes" (see Andreou, Pittis, and Spanos (1997) for an elaboration of this point).

Following in this line of research, Teichmoeller (1971) estimates the characteristic exponent (α)

for the daily returns from 30 NYSE stocks over the period from July 1962 to June 1967. He finds that the average characteristic exponent for the 30 stocks is less than two and fairly stable as the log price differencing interval is increased from one to ten days, indicating that the stock returns are generated by a stable distribution.

Barnea and Downes (1973) examine a sample of 81 securities, including only common stocks, and find that although α remains less than two, it does exhibit an upward trend as the differencing interval is increased. They hypothesize that these results may be caused by stock returns' being generated by either a mixture of non-Gaussian stable distributions with different scale parameters or a more complex mixture of distributions with finite variances.

The inconclusiveness of this research, in combination with resistance to the possibility that the distribution of securities returns could have infinite variance, has led a number of researchers to examine the hypothesis that returns are generated by a Gaussian model with independent increments, but one whose variance follows some stochastic process of its own (i.e., returns follow a subordinated stochastic process, or "mixture-of-normals" model). (Note that unlike the cross-sectional mixture of distributions that occurs across the stocks within a given portfolio, the "mixture-of-normals" referred to here is an intertemporal mixture across the different increments within the sequence of returns for one stock, so that the returns for that stock over different time intervals are assumed to follow normal distributions having constant means but variances that change with each time interval.) Such a process, while being conditionally normal, would have a marginal or unconditional distribution that is leptokurtic. Thus, such a process would have the favorable characteristics of independence and normality, but would also exhibit the "fat tails" so widely found in empirical studies.

Mandelbrot and Taylor (1967) show that a sequence with stable Paretian increments for which the characteristic exponent is less than two could be written in the form of a subordinated stochastic process with normal increments for which the variance follows a stable Paretian distribution of its own, with a characteristic exponent equal to $\alpha\text{-}1$ (implying that the expected value of the variance process is infinite, or undefined). Clark (1973) follows along similar lines to develop the "lognormal-normal" model. In this model, the variance follows a lognormal distribution, so that the variance has a finite mean, and as a result the marginal or unconditional distribution of returns has finite moments. However, this model has the disadvantage that the unconditional distribution has no closed form solution and must be expressed in integral form.

Blattberg and Gonedes (1974) argue that a better model is a subordinated stochastic process for which the variance follows an inverse gamma distribution. This leads to stock returns which follow the well-known Student's t distribution, a distribution which is completely described by a combination of three parameters—a mean parameter (μ), a scale parameter (σ), and the number of degrees of freedom (ν). This distribution is related to the stable Paretian distribution through the extremes of possible values for ν. If ν is zero, then the t-distribution is the symmetric stable Paretian distribution with α equal to one (the Cauchy distribution). As ν approaches infinity, the t-distribution tends toward the symmetric stable Paretian distribution with α equal to two (the normal distribution). Using likelihood ratio tests and examinations of distributional stability (the form of a stable Paretian distribution is stable regardless of the number of days' differencing used to calculate returns), Blattberg and Gonedes find that t-distributions with low (ten or less) degrees of freedom most adequately describe the returns of most of the 30 DJIA stocks.

Finally, Fielitz and Rozelle (1983) return to the possibility that stock returns follow a stable Paretian distribution and find evidence consistent with a mixture-of-distributions hypothesis. However, they are unable to distinguish between mixtures of normal distributions with changing variances and mixtures of nonnormal stable Paretian distributions with changing scale parameters. Furthermore, they obtain somewhat different results if the sequences of stock returns are randomized before the distribution parameters are estimated, indicating that the stock returns are not independent.

This result points out a major weakness that the above papers all have in common in testing their various distributional hypotheses—they all assume that securities returns follow a stochastic process with *independent* increments. Their only source of disagreement is on the exact nature of the probability distribution governing these independent increments. However, as Mandelbrot (1963) noted, and as Blattberg and Gonedes (1974) also found, securities returns appear to exhibit "volatility clustering." This clustering of variance indicates that, despite a lack of any significant autocorrelations among securities returns, they are not in fact independent; rather, they follow some sort of nonlinear dynamic process. Consequently, if, for example, stock returns follow a subordinated stochastic process for which the subordinated process has independent increments, then the variances for the subordinated process must be jointly, rather than independently, distributed across time. In the case of ARCH models (which were briefly described in Chapter 1 and will be discussed in more detail below), the subordinating process would be a deterministic moving average process for the variance of the subordinated process.

However, in addition to the ARCH class of models, there are literally an unlimited number of possible nonlinear dynamical models which could potentially describe the return generating processes for financial securities. Thus, the major statistical problem becomes that of detecting the existence of any nonlinear serial dependencies in the data, and then, given that such dependencies do exist, determining more precisely the exact form of the nonlinear process generating the returns.

2.2. Nonlinear Dynamics and the BDS Test

One of the most general and widely used tests for detecting nonlinear dependencies in a time series is the BDS test, which was developed by Brock, Deckert, and Scheinkman (1987). The BDS statistic has its origins in the correlation dimension plots of Grassberger and Procaccia (1983), which were developed for studying low-dimensional chaos in time series in physics applications. A chaotic dynamic process is a complex, but deterministic, nonlinear dynamic process. (A simple example is the logistic map $x_t = 4x_{t-1}(1-x_{t-1})$, where $x_t \in (0,1)$.) Such a process may look random, but is, at least in theory, potentially perfectly predictable. However, financial time series are more likely to follow nonlinear dynamic processes that are stochastic or random, rather than ones that are chaotic and deterministic. The BDS statistic was developed to enable researchers to detect the existence of any type of either of these two categories of nonlinear dynamics.

Under the null hypothesis that the increments of the time series in question are independent and identically distributed, the BDS statistic is as follows:

$$w_N(e,T) = \frac{\sqrt{T}[C_N(e,T) - C_1(e,T)^N]}{\sigma_N(e,T)} \xrightarrow{D} N(0,1) ,$$

where $C_N(e,T)$ is the sample correlation integral (an estimate of the probability that any two N-histories of the data are within e of each other, where N is the embedding dimension and e is the scaling parameter) and $\sigma_N(e,T)$ is the estimated standard deviation of the BDS statistic under the null hypothesis of independence.

It is important to note, though, that a rejection of the null hypothesis for the BDS test could mean any one or a combination of three major possibilities -- one, there are linear serial dependencies in the data; two, the time series is nonstationary; and three, there are nonlinear serial dependencies in the data, either chaotic or stochastic. So, a significant BDS statistic could signal the existence of an underlying nonlinear dynamic process. However, the BDS test will also be flagged by many other departures from the assumptions of i.i.d., including the existence of linear dependencies within the data. Thus, it would be absolutely crucial, if one were searching specifically for nonlinear dependencies, to first filter out the effects of any significant linear or structural dependencies (or regime shifts) in the data, such as autocorrelation, day-of-the-week effects, weekend effects, or month-of-the-year effects. After these effects have been filtered out in order to "pre-whiten" the time series data, then the BDS test can be used as a diagnostic test to determine whether any nonlinear dependencies remain within the data and must be modeled, and, subsequently, whether any such dependencies have been adequately modeled. Scheinkman and LeBaron (1989), Hsieh (1989), and Hsieh (1991) all use the BDS test for this purpose.

From a random sample of stocks on the CRSP data base, Scheinkman and LeBaron (1989) choose Abbott Laboratories for more thorough study, on the basis that its weekly stock returns look the most "random." After fitting an ARCH model to Abbott Laboratories daily stock returns, they use the estimated parameters to generate a simulated ARCH data set. Comparing this simulated ARCH data series to the original return data, they find that the simulated ARCH series actually look much more random than the original data. This implies that the true dynamic process underlying the stock returns is much more complicated than an ARCH process.

Hsieh (1989) examines the daily foreign exchange rates versus the U.S. Dollar for five major currencies -- the British Pound, the Canadian Dollar, the Deutsche Mark, the Japanese Yen, and the Swiss Franc. All five of these time series exhibit highly significant BDS statistics even after autocorrelation effects and linear holiday and day-of-the-week effects have been filtered out, thereby indicating the existence of strong nonlinear dependencies within these data series. He finds that a GARCH(1,1) model with either a Student's t or a generalized error distribution can describe the Canadian Dollar and the Swiss Franc exchange rates very well and the Deutsche Mark exchange rate reasonably well. While a GARCH(1,1) model can also account for most of the nonlinear serial dependencies within the British Pound and the Japanese Yen exchange rates, though, such GARCH models do not fit either of these time series very well. Similarly, Hsieh (1991) finds for value-weighted, size-decile portfolios of weekly stock returns from 1963 to 1987 that these returns also exhibit nonlinear serial dependencies and that conditional heteroskedasticity could be the source of these nonlinearities, but that none of the ARCH models seem to adequately describe these data.

In general, as was the case in the aforementioned studies, the BDS test has gained widespread usage in analyzing the residuals of GARCH-type models (e.g., Bollerslev, Engle, and Nelson (1994)). Brock, Deckert, Scheinkman, and LeBaron (1996) argue that this is due to the fact that the BDS test can be used as a misspecification test for temporal independence and linearity and that it is characterized by the invariance principle. But, though it is often used in this way, applying the BDS test to residuals from some preliminary estimation of a fitted model can be problematic. The resultant 'nuisance parameter' problem affects the behavior of the BDS statistic in finite samples and leads to the BDS test's having an actual size greater than its nominal size. This problem is exacerbated and exists even for large samples when the residuals come from a GARCH-type model, in which case the BDS test will lack power to reject a false model (see Brock, Hsieh, and LeBaron (1991)).

Given these caveats, while the BDS test can serve as a tool for analyzing nonlinearity in general, it functions best as a portmanteau test for the existence of dependencies in a given data set. Thus, if it is to be used, its most useful role is as a final specification check to determine whether the chosen econometric model adequately describes all the dependencies in the data. However, because it is a very general test which is sensitive to *any* type of dependency in the data, it is not a very useful tool for the initial examination and model selection phases of analysis. For these stages of analysis, it is more productive to examine the data using nonlinearity tests whose results can more readily be translated into specific categories of potential nonlinear models.

Hsieh (1989) divides the realm of nonlinear dependencies into two broad categories—additive nonlinear dependence and multiplicative nonlinear dependence. Additive nonlinearity, also known as nonlinearity-in-mean, enters a process through its mean or expected value, so that each element in the sequence can be expressed as the *sum* of a zero-mean random element and a nonlinear function of past elements. With multiplicative nonlinearity, or nonlinearity-in-variance, each element can be expressed as the *product* of a zero-mean random element and a nonlinear function of past elements, so that the nonlinearity affects the process through its variance.

2.3. Additive Nonlinearity

The paradigm of the normal linear time series model is the Autoregressive-Moving Average model of orders p and q, or more concisely the ARMA(p,q) model, given by:

$$y_t + \sum_{i=1}^{p} \phi_i y_{t-i} = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}$$

$$\varepsilon_t \sim NIID(0, \sigma_\varepsilon^2)$$

In the more general case in which normality is not assumed, ε_t is simply a zero-mean white noise (uncorrelated) process with constant variance. Expressing the ARMA model in vector notation yields:

$$y_t + \underline{\phi}^T \underline{y}_{t-1} = \varepsilon_t + \underline{\theta}^T \underline{\varepsilon}_{t-1}$$

The essential characteristic for this to describe a stable linear model is that the coefficients of the vectors $\underline{\phi}$ and $\underline{\theta}$ must be constant.

In the case of additive nonlinearity, however, the coefficients on the ARMA model are not constant, but instead vary systematically as a function of past data. A simple example of such systematic variation is the threshold autoregressive model of Tong and Lim (1980), in which the values of the ARMA parameters are step functions of the past data values. A broader and somewhat more widely used family of additively nonlinear models are bilinear models (cf. Granger and Andersen (1978) and Subba Rao and Gabr (1980)). A bilinear time series is one whose dynamics can be expressed in the form:

$$y_t + \underline{a}^T \underline{y}_{t-1} = \varepsilon_t + \underline{b}^T \underline{\varepsilon}_{t-1} + \underline{y}_{t-1}^T C \underline{\varepsilon}_{t-1} \ ,$$

where $\underline{a}$ and $\underline{b}$ are vectors of constant parameters and C is a matrix of constant parameters. A bilinear model could also be written in either of the following two forms:

$$y_t + (\underline{a} - C\underline{\varepsilon}_{t-1})^T \underline{y}_{t-1} = \varepsilon_t + \underline{b}^T \underline{\varepsilon}_{t-1}$$

$$y_t + \underline{a}^T \underline{y}_{t-1} = \varepsilon_t + (\underline{b} + C^T \underline{y}_{t-1})^T \underline{\varepsilon}_{t-1}$$

Thus, it is readily apparent that a bilinear model can also be viewed as an ARMA model with time-varying parameters whose values are linearly dependent upon previous realizations of the time series.

Under the assumption that the time series $\{y_t\}$ is third-order stationary, then the bilinear model could alternatively be written as:

$$y_t = \varepsilon_t + \underline{l}^T \underline{\varepsilon}_{t-1} + \underline{\varepsilon}_{t-1}^T Q \underline{\varepsilon}_{t-1}$$

This results in a more generalized version of the quadratic model of Hinich and Patterson (1985b). Conversely, if the time series is invertible (a property which, when present, facilitates forecasting), then its dynamics can be expressed as follows:

$$y_t = \underline{a}^T \underline{y}_{t-1} + \underline{y}_{t-1}^T B \underline{y}_{t-1} + \varepsilon_t$$

In this case, the matrix B consists of the bicorrelations of the process, which are equal to the bicovariances of the process scaled by the cube of the unconditional or marginal standard deviation for the process.

The bicovariances of a process are defined as:

$$E[(y_t - \mu_y)(y_{t-m} - \mu_y)(y_{t-n} - \mu_y)],$$

where m and n are greater than zero. If a time series is generated by a linear, Gaussian process, then its bicovariances should all be equal to zero. Conversely, if a process generating a time series has nonzero bicovariances, then the process in nonlinear and, even if the unconditional autocovariances of the process are all equal to zero, future elements of the series can be predicted via the *conditional* autocovariances. This is due to the fact that , for n greater than m,

$$E(y_t y_{t-m} y_{t-n}) = E[y_{t-n} E(y_t y_{t-m} | y_{t-n})] \neq 0.$$

In other words (assuming for simplicity that $\mu_y = 0$), the bicovariance can be seen to be the covariance of y_{t-n} with the condititional autocovariance between y_t and y_{t-m}, conditional on y_{t-n}. Thus, if a process has nonzero bicovariances, then this implies that the conditional autocovariances of the process are not constant, but are themselves random variables that can, in turn, covary with past returns.

The fact that linear processes have no nonzero bicovariances is the property underlying Hinich's (1982) bispectrum test. Assuming that $\{y_t\}$ is a third-order stationary, zero mean time series, then its bispectrum $B_y(f_1 f_2)$ is defined as the double Fourier transform of its bicovariance function:

$$B_y(f_1 f_2) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} E(y_t y_{t-m} y_{t-n}) e^{-2\pi i (f_1 m + f_2 n)}$$

It is a complex valued, spatially periodic function with a principal domain in the triangular set:

$$\Omega = \{0 < f_1 < 1/2, 0 < f_2 < f_1, 2f_1 + f_2 < 1\}$$

Conversely, the bicovariances can be expressed as the double inverse Fourier transform of the bispectrum:

$$E(y_t y_{t-m} y_{t-n}) = \iint_{\Omega} B_y(f_1, f_2) e^{2\pi i (f_1 m + f_2 n)} df_1 df_2$$

If the process $\{y_t\}$ is linear as well as stationary, then it can be expressed as:

$$y_t = \sum_{n=0}^{\infty} a_n \varepsilon_{t-n},$$

where $\{\varepsilon_t\}$ is a purely random, zero mean process and the a_n are constants. In this case, the bispectrum can be rewritten as:

$$B_y(f_1, f_2) = \mu_3 A(f_1) A(f_2) A^*(f_1 + f_2)$$

where $\mu_3 = E(\varepsilon_t^3)$, $A(f) = \sum_{n=0}^{\infty} a_n e^{-2\pi i f_n}$, and $A^*(f)$ is its complex conjugate. Since the spectrum of $\{y_t\}$ is:

$$S_y(f) = \sigma_\varepsilon^2 |A(f)|^2,$$

then whenever $\{y_t\}$ is linear the "squared skewness function" for $\{y_t\}$ is given by:

$$\Psi^2(f_1, f_2) \equiv \frac{|B_y(f_1, f_2)|^2}{S_y(f_1) S_y(f_2) S_y(f_1 + f_2)} = \frac{\mu_3^2}{\sigma_\varepsilon^6},$$

which is constant for all f_1 and f_2 in Ω. Thus, a stationary linear process will have a constant skewness function. Furthermore, if the $\{\varepsilon_t\}$ are also normally distributed, then μ_3 will equal zero, and the constant value that the skewness function equals will be zero. Ashley, Patterson, and Hinich (1986) show that this test is suitable for the detection of nonlinear dependencies in a time series regardless of whether it has been linearly pre-filtered. Thus, this test can be applied to either a pre-whitened data series or the original raw data set.

Hinich and Patterson (1985a) apply this test to stock market data. They examine the sample bispectra of the linearly pre-whitened daily returns from July 1963 through December 1977 (for a total of 3,881 observations for each stock) for fifteen stocks randomly selected from the New York and American Stock Exchanges. They find that both Gaussianity and linearity are strongly rejected for all fifteen of the stocks.

This implies that nonzero bicovariances are a significant feature of stock returns, a feature which Hsieh (1989) notes has been found by other researchers looking at stock returns for time periods during the 1980's. However, he does not find evidence that nonzero bicovariances play a significant role in exchange rate dynamics.

The previous tests, as well as many of the tests discussed in this chapter, rest on an implicit assumption that the form of any serial dependencies within a data series remains stable. To examine this assumption, Hinich and Patterson (1996) develop a bicorrelation portmanteau test that can be used to detect "epochs" of transient nonlinear dependencies within a data series. An autocorrelation portmanteau test similar to the Box-Pierce Q-statistic is also developed for the detection of transient linear dependencies. In applying these tests, the full sample is broken down into smaller windows of data, each of which is then tested for significant autocorrelations and bicorrelations. The bicorrelation portmanteau test, which is a time domain analog of the bispectrum test, uses what is called the "*H*" statistic to test the significance of the bicorrelations within a given window. Under the null hypothesis that the observations are i.i.d., this statistic is distributed as follows:

$$H_N = L^{-1}\sum_{s=2}^{L}\sum_{r=1}^{s-1}[G^2(r,s)-1] \xrightarrow{D} N(0,1)$$

where:

$$L = N^c,\ 0 < c < 0.5$$

$$G(r,s) = (N-s)^{-0.5}\sum_{k=1}^{N-s}u(t_k)u(t_k+r)u(t_k+s)$$

The authors apply this test to intra-day trade rates of return for 15 of the 30 DJIA stocks for a sample period from January 2, 1980 to August 30, 1985. They find that the overall sample exhibits significant linear and nonlinear dependencies. However, after breaking the sample down into smaller windows, they find that there are a number of short (two week long) pockets of nonlinear activity which seem to be driving these results. Outside of these windows of time, the nonlinearities are insignificant. This indicates that, in addition to nonzero bicovariances, nonstationarity may also play a major role in describing stock return dynamics.

The authors then extend this test procedure to search for deeper forms of transient nonlinear dependencies. In this extended procedure, the original data series is first dichotomized by setting all nonnegative returns equal to 1 and all negative returns equal to -1. Such a procedure would remove the effects of multiplicative nonlinearities, such as GARCH effects (to be discussed in the next section). This transformed data series is then divided into windows and examined via the autocorrelation and bicorrelation portmanteau tests. This test procedure is applied to intra-daily stock returns for eight NYSE stocks. Each of these sets of returns is found to exhibit long periods of time during which there is no evidence of linear or nonlinear dependencies interspersed with infrequent bursts of highly significant serial dependencies. Although the bursts of significant nonlinear activity appear only infrequently, however, they appear much more frequently, given the nominal size of the tests, than should be the case if the time series under study consisted of IID increments from a well-behaved distribution. Thus, these results provide further evidence that the return generating processes, even for relatively liquid, heavily traded stocks of NYSE-listed firms, may be characterized by both nonlinearity and nonstationarity.

2.4. Multiplicative Nonlinearity

The above results imply that bicovariances, hence additive nonlinearities, play an important role in describing stock returns, although the significance of this role appears to vary over time. However, multiplicative nonlinearity, or nonlinearity-in-variance, has received a far greater level of attention in the financial literature. This is a result of the successful development of the AutoRegressive Conditionally Heteroskedastic (ARCH) family of models (see Bollerslev, Chou, and Kroner (1992) for an extensive survey of the development and application of these models; see also Bollerslev, Engle, and Nelson (1994)), which is designed specifically to capture nonlinearities that enter a process through its variance.

Engle (1982) originated the ARCH class of models with his development of the ARCH(q) model,

$$
\begin{aligned}
&y_t = \varepsilon_t, \\
&\varepsilon_t \sim N(0,h_t) \\
&h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 = \omega + \alpha(L)\varepsilon_t^2,
\end{aligned}
$$

where $\omega > 0$ and $\alpha_i \geq 0$, and L denotes the lag operator. Such a process has the favorable characteristic that it can generate clusters of volatility such as have been observed in actual financial time series.

Bollerslev (1986) generalizes the order of the ARCH(q) model into the Generalized AutoRegressive Conditional Heteroskedasticity model of orders p and q, the GARCH(p,q) model:

$$
\begin{aligned}
&y_t = \varepsilon_t \\
&\varepsilon_t \sim N(0, h_t) \\
&h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t
\end{aligned}
$$

This process is covariance stationary if and only if $\alpha(L)+\beta(L)<1$, in which case the model could also be written as an infinite order ARCH model. Bollerslev (1988) proceeds to show that the

GARCH variance formulation can be rewritten in the form of a linear ARMA model for ε_t^2:

$$\varepsilon_t^2 = \omega + [\alpha(L) + \beta(L)]\varepsilon_t^2 - \beta(L)(\varepsilon_t^2 - h_t) + (\varepsilon_t^2 - h_t),$$

in which $(\varepsilon_t^2 - h_t)$ would be a white noise process. As this analogy suggests, the autocorrelation and partial autocorrelation functions for ε_t^2 can then be used to determine the orders for p and q. In most cases, though, the orders $p=q=1$ are found to suffice.

After finding that the GARCH(p,q) model accounted for some, but not all, of the leptokurtosis of the financial time series to which it had been applied, Bollerslev (1987) developed the Student's t GARCH, or GARCH(p,q)-t, model:

$$\begin{aligned} &y_t = \varepsilon_t \\ &\varepsilon_t \sim t_n(0, h_t) \\ &h_t = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t \end{aligned}$$

where n is the number of degrees of freedom for the conditional Student's t distribution. For finite n, this process will exhibit leptokurtosis not only in the original data series, but also in the series of standardized residuals. It is worth noting that this is the model which Hsieh (1989) found to adequately describe the Canadian Dollar and Swiss Franc foreign exchange rate series.

Spanos (1993) criticizes Bollerslev's GARCH-t model by arguing that its conditional distribution does not describe the conditional t-distribution in the case where the joint distribution is multivariate Student's t (see also Andreou, Pittis, and Spanos (1997)). As an alternative, Spanos utilizes the conditional distributional properties of the multivariate Student's t distribution to develop the Student's t AutoRegressive model with dynamic heteroskedasticity, the STAR($l,p;n$) model. The STAR model takes the form:

$$\begin{aligned} &y_t = \beta_0 + \sum_{i=1}^{l} \beta_i y_{t-i} + \varepsilon_t, \\ &\varepsilon_t \sim t_n(0, \omega_t^2) \\ &\omega_t^2 = \left[\frac{n}{n+t-3}\right]\sigma^2\left[1 + \sum_{i=1}^{t-1}\sum_{j=-p}^{p} q_{|j|}(y_{t-i} - \mu)(y_{t-i-j} - \mu)\right] \end{aligned}$$

where $\varepsilon_t \equiv y_t - E(y_t | \Im_{t-1})$, $\mu \equiv E(y_t)$, and $n>2$ is the number of degrees of freedom for the multivariate Student's t distribution. In this model, the conditional mean is linear in the conditioning variables, while the conditional variance is a quadratic function of all past conditioning information but is parameterized with only $p+1$ unknown q_j's, so that it is equivalent to a sequentially smoothed version of the unconditional variance. McGuirk, Robertson, and Spanos (1993) apply this model to four of the five foreign exchange rate series examined by Hsieh (1989), including the Japanese Yen, the German Mark, the Swiss Franc, and the British Pound. McGuirk, et. al., find that GARCH models do not fit these data well, but that STAR models can be fit that adequately describe the statistical properties for each of these four series.

The successes of these conditionally heteroskedastic models leads to the question of how nonlinear dependencies in the form of conditional heteroskedasticity within a time series can initially be detected. Two related tests for the detection of such nonlinearities are developed by McLeod and Li (1983) and Engle (1982). McLeod and Li show that the Box-Pierce Q-statistics of the squared residuals of an ARMA model can be used to test for nonlinear dependence. As Hsieh (1989) notes, this method can also be applied to the unfiltered time series. Alternatively, Engle (1982) uses a Lagrange multiplier test for which squared residuals are regressed on lagged squared residuals. Thus, these tests are related in that the Engle test examines the partial autocorrelation function of the squared residuals, while the McLeod and Li test utilizes the autocorrelation function. As was noted previously, if conditional heteroskedasticity is present, then simultaneous examination of both of these functions can be used to determine the orders of a GARCH model.

However, there is a problem with using this categorization system in that additive and multiplicative nonlinearity can mimic each other to some extent, and both of these types of nonlinearity can generate sequences for which the squared returns are autocorrelated. Thus, a GARCH process may not be the true process underlying the observed squared-return autocorrelation function. In fact, as Weiss (1986) points out, the McLeod and Li test can serve as a diagnostic for detecting the orders of either a GARCH-type process or a bilinear process. Furthermore, some types of nonlinearity could even resemble nonstationary linear processes. Data generated by an ARCH(q) process, for example, are observationally equivalent to data generated by a time-varying-parameter MA(q) process. Thus, a significant McLeod and Li test statistic could indicate the existence of multiplicative nonlinear serial dependencies in the form of conditional heteroskedasticity, additive nonlinear serial dependencies in the form of bilinearity, or nonstationarity in the form of linear dependencies with time-varying parameters, or some combination of these three possibilities.

2.5. Volatility Persistence and Nonlinear Dependence

Despite the difficulties in distinguishing between GARCH-type processes and other nonlinear or nonstationary stochastic processes, GARCH-type models have been ubiquitous in empirical finance research. An important factor behind their widespread usage is the direct, linear relationship between the variance of a financial time series and its expected value that is posited by many theories in finance, most notably the Sharpe-Lintner CAPM. Because GARCH-type models provide an explicit functional form to describe the evolution over time of the variance of a time series of returns, such models provide a ready means to empirically examine the relationship between “risk” and “return” on financial markets, so long as the models chosen for such study are statistically well-specified for the specific time series in question.

One attempt to directly estimate this relationship in one step was developed by Engle, Lilien, and Robins (1987), who originated the GARCH-in-mean, or GARCH-M model:

$$y_t = \delta h_t + \varepsilon_t$$
$$\varepsilon_t \sim N(0, h_t)$$
$$h_t = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)h_t$$

for which δ represents the coefficient of investors’ aggregate risk aversion. An interesting feature of this model, related to the above discussion, is that it exhibits nonlinear dependencies in both its mean and its variance terms. As such, it “bridges the gap” of the previous two sections of this chapter.

Unfortunately, apart from the empirical question of whether such a model would provide a statistically adequate fit to a given financial time series, French, Schwert, and Stambaugh (1987) note that this model suffers from a significant theoretical shortcoming. Although the relationship the model describes pertains to ex ante expected returns, the model must be fitted to ex post realized returns. And if the hypothesized relationship is correct, an increase in the variance of returns will lead to an increase in required expected returns. But, if the future expected payoffs remain unchanged, then for future returns to be commensurate with the level of variance, current prices must fall, causing negative current returns. In other words, a positive relationship between expected returns and variance can lead to a negative relationship between variance and ex-post realized returns. Thus, when fitting the model to a time series, the resulting estimates of δ, the measure of investors' aggregate level of risk aversion, must be interpreted with caution.

Moving away from these difficulties, an alternative source of potentially useful information for examining the risk-return tradeoff is provided by even the more basic GARCH-type models. This is the measure of "volatility persistence" that can be computed from the parameters of the estimated variance equation. For the variance of a GARCH-type process to be stationary, the parameters in the variance equation, excluding the intercept term, must sum to less than one (i.e., in general we must have $\alpha(L)+\beta(L)<1$ for covariance stationarity; for the GARCH(1,1) model, this reduces to $\alpha+\beta<1$). The closer is the sum of these variables to one, the less stable the variance will be in the long run, and the more permanent will be changes in the level of volatility as a consequence of "volatility shocks." Conversely, the smaller is this sum relative to one, the more transient will be the effects of the volatility shocks, and the less of an adjustment there will be to expected returns.

One of the earliest applications of GARCH models along these lines was as to study the causes of the U.S. stock market collapse of the mid-1970's, during which the market declined by 45%. Chou (1988) used a GARCH(1,1) model to analyze weekly NYSE value-weighted index data from CRSP for the period from 1962 to 1985 and found that, while the estimated GARCH parameters α and β were such that $\alpha+\beta<1$, their sum was sufficiently close to one that the hypothesis of a unit root in the variance equation could not be rejected. This possibility would mean that volatility shocks would have near-permanent effects and that a better model for the data may have been the Integrated GARCH, or IGARCH, model of Engle and Bollerslev (1986), for which the variance of the process follows a random walk and volatility shocks have infinite lives.

Many other researchers, examining somewhat different data sets, also obtain findings which are consistent with the possibility that volatility shocks have infinite lives. French, Schwert, and Stambaugh (1987) find evidence implying the possibility of a unit root in the variance of the S&P 500 daily index returns. Additionally, they find a negative relationship between stock market returns and the unpredicted component of the variance of these returns. Pagan and Schwert (1990) find evidence of a unit root in the variance of U.S. stocks. The possibility of a unit root also cannot be rejected for the variance of the monthly size-ranked stock portfolios studied by

Schwert and Seguin (1990). Finally, using Exponential GARCH, or EGARCH, a GARCH formulation in which $\ln\varepsilon_t^2$ rather than ε_t^2 follows an ARMA process, Nelson (1989 and 1990) still finds evidence supporting a unit root.

However, Harvey, Ruiz, and Sentana (1992), upon re-estimating foreign exchange rate data originally examined by Diebold and Nerlove (1989), find that the volatility persistence parameter drops from 0.97 to 0.93 when the assumption of conditional normality is changed to that of a conditional t-distribution. Thus, while the estimated level of volatility persistence may not be sensitive to the form of the GARCH variance equation, it is sensitive to the choice of the conditional distribution.

But not only is the level of volatility persistence sensitive to the choice of conditional distribution, it is also sensitive to the choice of the *joint* distribution. As Spanos (1993) and McGuirk, Robertson, and Spanos (1993) note, if the joint distribution of the return series is multivariate Student's t, then the proper model specification is the STAR model, for which the memory of the conditional variance is as long as the sample. Given this long memory, if the STAR model is indeed the better specification for financial time series, then the finding of high levels of volatility persistence in fitted GARCH models is not surprising, and, conversely, such a finding in fitted GARCH models may be indicative that STAR rather than GARCH is the more appropriate specification for the data. In general, while both types of models exhibit dynamic heteroskedasticity, the conditional variance for the STAR model is smoother, tending toward a constant limit and representing a long-run relationship, while the conditional variance for the GARCH or GARCH-t models is more volatile, more closely capturing short-term fluctuations. But if STAR is the appropriate model specification, then the effectively infinite memory of the STAR model renders moot the question of volatility persistence.