THE JOURNAL OF FINANCE • VOL. XLIV, NO. 4 • SEPTEMBER 1989

Nonnormalities and Tests of Asset Pricing Theories

JOHN AFFLECK-GRAVES and BILL MCDONALD*

ABSTRACT

The robustness of the multivariate test of Gibbons, Ross, and Shanken (1986) to nonnormalities in the residual covariance matrix is examined. After considering the relative performance of various tests of normality, simulation techniques are used to determine the effects of nonnormalities on the multivariate test. It is found that, where the sample nonnormalities are severe, the size and/or power of the test can be seriously misstated. However, it is also shown that these extreme sample values may overestimate the population parameters. Hence, we conclude that the multivariate test is reasonably robust with respect to typical levels of nonnormality.

IN TRADITIONAL HYPOTHESIS TESTING, a nonrandom test maps the values of a random variable into a sample space dichotomized into regions where a hypothesis is either accepted or rejected. There are three possible outcomes from this process: (1) a correct decision, (2) a false rejection (Type I error), or (3) failure to reject the hypothesis when it is false (Type II error). Of the latter two types of error, an error of the first kind is usually considered less desirable. So typically, a level of significance is selected with low probability of Type I error (e.g., 0.05 or 0.01), and a test is chosen so as to maximize power (minimize probability of Type II error) for the specified level of Type I error.[1] Knowledge of the relative level of these two errors is critical in assimilating the results of an experiment. For example, if the power of a test is equal to its significance level (i.e., a weak test), rejection of the null contains zero information.[2] Additionally, in constructing parametric tests of hypotheses, it is necessary to assume some distribution for the underlying data. Consequently, when using parametric tests, rejection of the null is only equivalent to rejection of at least one of the underlying hypotheses (i.e., the null hypothesis or the distributional assumption).

Interestingly, the size (significance level) and power of procedures used to test

* Both authors from College of Business, University of Notre Dame. This research was funded in part by the Herrick Foundation Fund. Data were provided by the Jesse H. Jones Research Data Base. We are grateful to Jay Shanken, Rich Sheehan, and workshop participants at the University of Notre Dame for helpful comments. The manuscript has also benefited from the substantive comments of an anonymous reviewer.

[1] Some authors distinguish between the terms "size of test" and "level of significance." (e.g., Lehmann (1986, pp. 68–71)). Others (e.g., Kendall and Stuart (1967)) prefer to avoid the rhetoric of such words as "significance." We will use these terms synonymously throughout the paper.

[2] Selection of the significance level is in fact somewhat arbitrary. Shanken (1987b) considers the choice of significance level for the GRS test from a Bayesian perspective. For more general procedures where the level is selected relative to power, see Arrow (1960). Burgstahler (1987) presents a discussion of the impact of the power/size relationship on posterior beliefs.

asset pricing models have received relatively little attention. Historically, this may be attributable to the sequential design of Fama-MacBeth (1973)-type methodologies, where it is difficult to assess accurately the true power and size of the test. In this paper we focus on a test of asset pricing that resolves a number of previous econometric issues; however, it also requires multivariate normality of the market model residuals.

A multivariate test of asset pricing proposed by Gibbons, Ross, and Shanken (1986) (henceforth GRS) has recently received considerable attention in the literature because of its rich interpretive features (e.g., Kandel and Stambaugh (1987) or Shanken (1987a)). Under ideal conditions the power of the GRS test can be easily determined, and it is the uniformly most powerful invariant test. Both GRS and MacKinlay (1987) extensively examine the power of this test under the assumption of multivariate normality of the market model residuals. However, relatively little is known about the effects of violating the multivariate normality assumption in the context of the GRS test.

The purpose of this paper is to evaluate the effects of nonnormality on the size and power of the GRS test. We first carefully re-examine the degree of nonnormality present in monthly data over a series of five-year intervals using alternative normality tests that have been shown to have desirable properties. Next, a method of simulating the first four marginal moments of the multivariate market model residuals is presented and is used to examine the effects of nonnormalities. We find that, for the typical levels of nonnormalities measured in market model residuals, the GRS test is robust. However, for nontrivial deviations from normality observable in some periods, both the size and power of the test can be seriously misstated (in one case by a factor greater than two). Subsequent analysis of random samples of portfolio returns indicates that such extreme deviation from normality is unusual (i.e., it is most likely a sampling phenomenon). Thus, the GRS test appears to be reasonably well specified in tests of asset pricing.

The paper is organized as follows. In the next section the GRS test is reviewed in the context of this paper. In Section II we briefly review previous studies of the distributional properties of security returns and, separately, the properties of various tests of normality. Section III describes the data and simulation design. Results of the normality tests for market model residuals and the robustness tests for the GRS statistic are presented in Section IV. Conclusions are provided in the final section.

I. The GRS Test

Gibbons, Ross, and Shanken (1986) propose a test for the efficiency of a given portfolio based on the residuals and estimated intercepts obtained from a regression of the individual security (or portfolio) returns on the return of a given portfolio (i.e., a market model regression where the specified portfolio is assumed to be a surrogate for the market). More specifically, given a fixed riskless rate of interest for each time period, GRS assume that the following model is well specified:

$$\tilde{r}_{it} = \alpha_{ip} + \beta_{ip}\tilde{r}_{pt} + \tilde{e}_{it} \qquad \text{for } i = 1, 2, \cdots, N; \quad t = 1, 2, \cdots, T, \tag{1}$$

where

$\tilde{r}_{it}$ = the excess return on asset i in period t,
$\tilde{r}_{pt}$ = the excess return on portfolio p, whose efficiency is being tested,
N = the number of assets being tested,
T = the number of time-series observations on returns, and
$\tilde{e}_{it}$ = the disturbance term for asset i in period t.

The asset returns are assumed to be independent and identically multivariate normally distributed through time with

$$E(e_t) = 0$$

and

$$E(e_t, e_s) = \begin{cases} \Sigma, & t = s, \\ 0, & t \neq s, \end{cases}$$

where Σ is the $N \times N$ disturbance covariance matrix.

GRS show that a necessary condition for portfolio p to be mean-variance efficient is that $\alpha_{ip} = 0$ for all $i(i = 1, 2, \cdots, N)$. To test this condition, they propose the test statistic:

$$\Gamma = [T/(T-2)][(T-N-1)/N](1 + \hat{\theta}^2_p)^{-1}\hat{\alpha}_p'\hat{\Sigma}^{-1}\hat{\alpha}_p, \tag{2}$$

where

$\hat{\alpha}_p' = (\hat{\alpha}_{1p}, \hat{\alpha}_{2p}, \cdots, \hat{\alpha}_{Np})$,
$\hat{\alpha}_{ip}$ = the ordinary least squares estimate of α_{ip} in equation (1),
$\hat{\theta}_p = \bar{r}_p/s_p$,
$\bar{r}_p$ = the sample mean of $\tilde{r}_{pt}$,
s^2_p = the maximum likelihood estimate of the variance of $\tilde{r}_{pt}$, and
$\hat{\Sigma}$ = the maximum likelihood estimate of Σ, the variance-covariance matrix of $\tilde{e}_{it}$.

Under the null hypothesis and assuming that the error terms are multivariate normal, Γ has a central F-distribution with N and $T - N - 1$ degrees of freedom. (See Anderson (1984).) The primary objective of this paper is to examine the effect that nonnormalities in the market model residuals may have on the significance level and power of the GRS test.

II. Tests of Normality

Throughout the paper we use the marginal distributions to test and simulate the multivariate distribution. Although marginal normality does not necessarily imply multivariate normality, the presence of nonnormalities is often reflected in the marginal distributions. (See Mardia (1980) and Gnanadesikan (1977) for a discussion of this issue and a thorough review of recent advances in tests of normality.) Strict multivariate tests of normality (see, for example, Mardia (1980)) are possible; however, our simulation techniques can only mimic the

marginal moments of the joint distribution. Thus, we focus on the marginal moments.

We begin with a very brief literature review of normality tests for security returns. This is followed by a description of the normality tests used in this study.

A. Previous Research on Security Returns

Many previous studies have considered the distribution of price changes in speculative series. Early work (e.g., Bachelier (1964)) suggested that the distribution should be normal; later, however, Mandelbrot (1963) and others provided evidence rejecting normality and indicating that the price change distributions for many speculative series were leptokurtic. This led to the suggestion that the underlying distribution of stock returns was stable Paretian.

Subsequently, most studies of the distribution of stock returns have assumed nonnormality and focused on whether the distribution is better represented by the stable Paretian family (e.g., Fama and Roll (1971)) or by a subordinated stochastic process generated from a mixture or combination of distributions (e.g., Clark (1973)). Similar evidence of nonnormal behavior has been found in the distribution of the residuals from the market model (e.g., Blume (1968)), despite the fact that only estimated residuals from an OLS (or other) regression can be observed, resulting in the so-called "supernormality" problem. (See Gnanadesikan (1977).) In addition, Bookstaber and McDonald (1987) have suggested that the generalized beta distribution of the second kind may be a more appropriate family of distributions for stock returns.

Although some derivations in asset pricing theory rely on normality of returns, it is normality of the market model residuals that is critical in both Fama-MacBeth-type tests and multivariate tests of the theories. While nonnormality of the underlying return data does not necessarily imply nonnormality of the market model residuals, it nevertheless suggests that the distribution of the residuals should be carefully examined before making distributional assumptions. Despite the concerns about normality raised by these previous studies, most studies of asset pricing theories have either explicitly or implicitly assumed normality of the residuals in the market model regression (e.g., Fama and MacBeth (1973) or Shanken (1985)). A possible explanation for this is that many asset pricing tests have used monthly data, and it has been argued that the normal distribution provides a good working approximation for monthly returns (Fama (1976)). The objective of this paper is to demonstrate that even monthly residuals from the market model can display severe departures from normality for both individual securities and portfolios and, more importantly, to examine the effect of such departures on a particular asset pricing test.

B. Tests of Normality

Four different tests of normality are used in this study: the Studentized range (SR), the Shapiro-Francia (1972) test (W'), the sample ratio test for skewness ($\sqrt{b_1}$), and the sample ratio test for kurtosis (b_2). As previously discussed, these

univariate tests will be used to evaluate the marginal distributions of the multivariate sample.

The Studentized range (see David, Hartley, and Pearson (1954)) is defined as

$$SR = (x_{(T)} - x_{(1)})/[\Sigma(x_i - \bar{x})^2/(T - 1)]^{1/2}, \tag{3}$$

where $x_{(T)}$ is the Tth order statistic (highest) and $x_{(1)}$ is the first order statistic (lowest). This statistic has been widely used as a test for normality in speculative price series (e.g., Fama and Roll (1971) and Fama (1976)). Unfortunately, as Shapiro, Wilk, and Chen (1968) demonstrate in their simulation study, while the SR provides an excellent test against symmetric alternatives, it has virtually no sensitivity to asymmetry. Most subsequent studies confirm this result, generally concluding that the SR is dominated by the Shapiro-Wilk (1965) W and Shapiro-Francia W' statistics. Consequently, the SR tests are presented in the first two tables of this study only to allow comparison with prior studies of security returns that have incorporated this test.

The Shapiro-Francia W' statistic (see Shapiro and Francia (1972)) is defined as

$$W' = (\Sigma b_i x_{(i)})^2/\Sigma(x_i - \bar{x})^2, \tag{4}$$

where $x_{(i)}$ is the ith order statistic and the b_i depend only on the expected values of the normal order statistics. The statistic is a modification of the Shapiro-Wilk W test for sample sizes greater than 50. This statistic is generally regarded as a good omnibus test of normality. (See D'Agostino and Pearson (1973).)

The earliest statistical tests for departures from normality were based on the sample skewness and sample kurtosis. In particular, tests were based on the distribution of the standardized third and fourth moments

$$\sqrt{b_1} = m_3/(m_2^{3/2}), \tag{5}$$

$$b_2 = m_4/m_2^2, \tag{6}$$

where

$$\begin{aligned} m_r &= \Sigma(x_i - m'_1)^r/T, \qquad \text{for } r > 1, \\ m'_1 &= \Sigma x_i/T, \text{ and} \\ x_1, \cdots, x_T &= \text{a random sample of size } T \text{ from a univariate normal population.} \end{aligned}$$

Because the statistics $\sqrt{b_1}$ and b_2 represent the sample skewness and kurtosis, respectively, these tests have intuitive appeal; i.e., they provide some insight into the nature of the deviation from normality. The exact distribution of these statistics under the null (of normality) is difficult to obtain, and approximations to the probability integrals must be used (e.g., D'Agostino and Pearson (1973)). Nevertheless, simulation studies indicate that these sample moment ratio tests are reasonably powerful, especially in cases where the direction of the deviation from normality is known. (See Pearson, D'Agostino, and Bowman (1977).) In cases where directional tests are not obvious, omnibus tests based on a combination of $\sqrt{b_1}$ and b_2 have been proposed (e.g., D'Agostino and Pearson (1973)). Unfortunately, as Bowman and Shenton (1975) observe, such omnibus tests are not fully justified because the statistics $\sqrt{b_1}$ and b_2 are not independent. In this

paper we will restrict our attention to the individual tests of $\sqrt{b_1}$ and b_2 and use the statistics primarily to gauge the extent of the deviation from normality in terms of the familiar concepts of skewness and kurtosis.

In summary, we initially consider four tests of normality: the Studentized range (SR), the Shapiro-Francia W', and the $\sqrt{b_1}$ and b_2 statistics. The SR test is included for comparison with previous studies. The W' is reported as an omnibus test that has been shown to be relatively powerful. After some subsequent empirical comparisons, we will focus on the $\sqrt{b_1}$ and b_2 statistics since they prove to dominate the SR test and are comparable to the omnibus W' test (in terms of number of rejections). Additionally, the skewness and kurtosis statistics provide a description of the nonnormalities. In the following section we outline the simulation design for the study.

III. The Simulation Design

To determine the effects of distributional aberrations on the size and power of the GRS test, it is necessary to simulate specific deviations from normality. In this section we describe a technique for generating nonnormal multivariate random variables and then outline the procedure used to evaluate the robustness of the GRS test. Before describing the simulation procedure, the data and sample structure used in the study are discussed.

A. Data and Sample Structure

Data for the study were taken from the CRSP monthly files.[3] Initial distributional tests of the market model residuals are performed for each of the five-year intervals from January 1931 to December 1985. To be included in the sample, the security must have complete return data for the sample period plus complete return data from the previous five years and the number of outstanding shares for the final month preceding the sample period. The latter two requirements provide for pre-period measures of beta and size subsequently used to form portfolios.

From the eleven time intervals, four were selected in order to facilitate the computational magnitude of the simulations. Three of the four, 1/36–12/40, 1/41–12/45, and 1/76–12/80, were selected because they exhibited relatively high levels of nonnormalities. One additional period, 1/81–12/85, was selected as a current period that exhibits relatively low levels of nonnormality.

B. Generating Multivariate Nonnormal Errors

One of the difficulties in testing the robustness of the GRS statistic is simulating nonnormal distributions that parallel their empirical counterparts. Fleishman (1978) presents a method for generating univariate nonnormal errors

[3] Much of the parameterization of our tests parallels MacKinlay (1987) (e.g., portfolios of twenty, size and beta sorting, etc.). Our focus on monthly data is consistent with previous tests of asset pricing. MacKinlay shows that there is little gain in increasing the sample interval to weekly data. (Also, obviously, any shorter interval can consider only post-1962 data using the CRSP data.)

with specified first four moments. His method is not based on the widely used inverse distribution function method (used for nonnormal univariate distributions such as the exponential) but utilizes knowledge of the first four moments of the power transformation, as in

$$Y = a + bX + cX^2 + dX^3, \tag{7}$$

where X is a standardized normal deviate and a, b, c, and d are constants that can be selected so that Y has the desired first four moments. By evaluating the first four moments of Y, a system of four equations with four unknowns can be used to determine a, b, c, and d. (See equations (5), (18), (19), and (20) in Fleishman.) Since the equations are nonlinear, they are solved using optimization methods.

Vale and Maurelli (1983) show how the Fleishman procedure can be modified to generate multivariate nonnormal variables with a specified covariance structure and specified first four marginal moments (i.e., a different specified mean, variance, skewness, and kurtosis for each marginal distribution). The method requires the ability to generate multivariate normal errors with a specified covariance structure and then applies a similar power transformation as in equation (7). If $r_{Y1,Y2}$ is the correlation between the nonnormal variables Y_1 and Y_2 (in our case the market model residuals) and $r_{X1,X2}$ is the corresponding correlation of the normal variates, then Vale and Maurelli show that solving the polynomial

$$r_{Y1,Y2} = r_{X1,X2}(b_1b_2 + 3b_1d_2 + 3d_1b_2 + 9d_1d_2) + r^2_{X1,X2}(2c_1c_2) + r^3_{X1,X2}(6d_1d_2) \tag{8}$$

for $r_{X1,X2}$ provides the correlation between X_1 and X_2 required to generate the desired post-transformation correlation $r_{Y1,Y2}$.[4]

C. Simulating the GRS Test Procedure

As the GRS test statistic, Γ in equation (2), requires the inversion of the estimated residual covariance matrix, N (the number of securities or portfolios used in the test) must be less than the number of observations T. (In fact, for positive degrees of freedom $N < T - 5$.) Consequently, most empirical applications of the GRS test have used portfolios rather than individual securities (e.g., MacKinlay (1987)). Because of stationarity considerations regarding the market model parameters, we limit our study to sixty-month periods and choose $N = 20$.[5] Thus, we examine the significance level and power of the GRS test using twenty portfolios. The GRS test is general enough to test a variety of asset

[4] It is worth noting two disadvantages of the Fleishman procedure as pointed out by Tadikamalla (1980). First, the exact distribution of Fleishman's random numbers is not known, as the method merely ensures a distribution with the same first four moments. In our case, we believe this to be an advantage as it avoids the problem of specifying an alternative distribution (both form—e.g., stable Paretian—and parameters). Second, the method cannot be used to generate variates with all possible combinations of skewness and kurtosis. Where such situations occurred in our simulations, the closest achievable combination was used.

[5] GRS show that "... when five years of monthly data are used, 20 to 30 assets may be appropriate" (p. 18). For sixty-month intervals, MacKinlay (1987) finds the nominal power to be insensitive to the choice of $N = 20$ or $N = 40$.

pricing specifications. To evaluate the power of the test we must consider a specific alternative. We use the Sharpe-Lintner model as the null hypothesis being tested against a specific alternative where the intercept is assumed to be different from zero by a given constant.

The following is an outline of the simulation design used to examine the effect of nonnormalities on the significance level of the GRS tests.

(a) Four nonoverlapping five-year periods were selected: 1936–1940, 1941–1945, 1976–1980, and 1981–1985. As previously explained, these are not a random selection but, rather, include the three periods that displayed the most evidence of nonnormality plus a recent period where nonnormalities are less dominant.

(b) For each period, all securities with available data were ranked on pre-period betas, and twenty beta-sorted portfolios were formed.

(c) For each of the twenty portfolios, the market model parameters α_i and β_i ($i = 1, 2, \cdots, 20$) were estimated, as were the sixty residuals e_{it} ($i = 1, 2, \cdots, 20$; $t = 1, 2, \cdots, 60$). The first four sample marginal moments of the residuals were computed for each portfolio, as was the cross-sectional covariance matrix of the estimated residuals. (As before, we assume that the residuals are serially independent.) For each period examined in the simulation, these initial estimates were fixed as the population moments and covariance matrix.

(d) Multivariate nonnormal errors were generated using the Vale and Maurelli (1983) method so as to provide a distribution with the same first four marginal moments for each portfolio and the same cross-sectional covariance structure.[6] That is, errors e^*_{it} were generated such that the ith set of sixty residuals came from a distribution with the same first four moments as portfolio i and such that the cross-sectional covariance structure was consistent with the original sample.[7]

(e) Simulated excess returns were then generated as

$$r^*_{it} = \alpha_i + \beta_i r_{pt} + e^*_{it},$$

where

α_i = 0
β_i = $\hat{\beta}_i$ estimated in step (c),
r_{pt} = the excess return on the CRSP equally weighted index in period t, and
e^*_{it} = the simulated error term from step (d).

(f) The simulated excess returns were then regressed against r_{pt} to obtain

[6] In this study, multivariate normal variates are generated using the IMSL subroutine GGNSM.

[7] The Vale and Maurelli (1983) method suffers from the same weakness as the Fleishman (1978) method. Consequently, it was not possible to obtain every required combination of skewness and kurtosis. When this occurred (only in the 1/36–12/40 beta-sorted and the 1/41–12/45 size-sorted cases), the largest values of skewness and kurtosis were moved toward the normal levels by a factor of ten percent (i.e., skewness was moved toward 0 and kurtosis was moved toward 3).

OLS estimates of α_i, β_i, e^*_{it}, and Σ. Using this data, the GRS statistic Γ (equation (2)) was computed.

(g) Steps (d) through (f) were repeated 5,000 times, and the number of times Γ fell above the central F-value with a nominal significance level of 0.05 (i.e., the number of rejections) was recorded.

The entire procedure was repeated for each of the four periods examined in this study. In addition, the process was repeated using size-sorted instead of beta-sorted portfolios. Note that, since $\alpha_i = 0$ in these tests, comparing the observed significance level to the nominal significance level provides a measure of misspecification in the size of the test.

Additionally, the process was used to examine the effect of nonnormalities on the power of the GRS test by setting $\alpha_i = (1 - \beta_i)\gamma$, where $\gamma = 0.008$ (i.e., ten percent per annum) in stage (e). Note, however, that, when this comparison is based on tests using the central F for the critical value, the difference in power is not readily apparent since the actual size of the test may differ from the nominal size as a result of the distributional misspecification. To address this issue, we repeat step (g) for the cases where $\alpha_i = (1 - \beta_i)\gamma$ using the 95th percentile of the simulated null distribution of Γ (i.e., the empirical distribution of Γ from step (g) when $\alpha_i = 0$) as the critical value rather than the five percent point of the central F distribution. While this does not provide an exact critical value (since the simulated null distribution is based on only 5,000 replications), it provides a reasonably good estimate with the advantage that the size of the test is maintained at approximately five percent.[8]

IV. Results

The results obtained indicate that, for both individual securities and portfolios, significant departures from normality are evident. This is true for a variety of normality tests including the Studentized range, the Shapiro-Francia W', and the sample moment ratio tests $\sqrt{b_1}$ (for skewness) and b_2 (for kurtosis). The results also indicate that the degree of nonnormality depends on the time period examined, with security returns in some periods displaying severe departures from normality while other periods display only minor departures from normality.

Having established that security returns and market model residuals both indicate evidence of nonnormalities, a simulation study is performed to ascertain the effects of such nonnormalities on the GRS test. For several nonoverlapping time periods, simulated residuals are generated from nonnormal distributions chosen to have the same first four marginal moments as actual portfolios in that time period and to have the same covariance structure across portfolios. This

[8] In the spirit of Gibbons and Shanken (1987), we can compute the standard error for the proportion of any random sample of size 5,000 that will lie above the sample 95th percentile ($P'_{0.95}$) as $\sqrt{(P(1 - P)/5{,}000)}$, where P is the true (population) probability that a sample value lies above $P'_{0.95}$. While P being a population parameter is unknown, we know that it should be of the order of five percent, which implies a standard error of approximately 0.003. More conservatively, the standard error is maximized at $P = 1/2$ and, hence, is certainly less than 0.007.

provides evidence of how nonnormalities actually inherent in the securities data might affect the GRS test.

The results indicate that, for small deviations from normality, the GRS test is reasonably robust. However, for more serious deviations from normality (e.g., in the period 1936–1940), the significance level of the GRS test assuming normality substantially understates the actual significance level (by as much as a factor of two). In other words, using the GRS test under the assumption of normality and choosing a critical point from the central F can result in a test with probability of rejection under the null significantly greater than the specified significance level. Similarly, under a given alternative hypothesis, the power of the test (i.e., probability of rejecting the null) may be significantly higher than the theoretical power calculated assuming normality if the underlying distribution is nonnormal. Thus, nonnormality in the underlying residuals can lead to a significantly higher significance level and power than that implied in the choice of the critical F under the assumption of normality. Subsequent tests, however, indicate that the level of nonnormality necessary to produce a substantive misspecification of the test is not typical in random samples of security data.

A. Tests of Normality

We begin by replicating the results of Fama (1976). Specifically, we examine the distribution of monthly returns over the period January 1951 to June 1968 for the thirty Dow Jones Industrial stocks. The results are summarized in Table I, which presents the Studentized range as reported by Fama (SR_F) together with the studentized range (SR) calculated using the current CRSP files. As can be seen, our results are almost a perfect replication of Fama's.[9] In addition, Table I presents tests of normality using the Shapiro-Francia (W') test and the directional tests $\sqrt{b_1}$ and b_2. As expected (given the simulation results of, for example, Pearson, D'Agostino, and Bowman (1977)), both the Shapiro-Francia and the directional tests result in a greater number of securities for which the null hypothesis of normality is rejected than does the studentized range.

Two minor points are worth noting. First, the directional tests result in more rejections than the Shapiro-Francia test when testing individual monthly returns for normality. This is also apparent when using the omnibus rectangular test of Pearson, D'Agostino, and Bowman (1977) assuming independence between $\sqrt{b_1}$ and b_2. Second, only six of the thirty securities examined display consistency with the normality assumption at the ten percent level for both the $\sqrt{b_1}$ and b_2 directional tests.

We conclude that the results presented in Table I indicate that nonnormality is evident in the monthly return series for individual securities. Moreover, the use of more powerful statistical tests indicates that such nonnormalities may be even more prevalent than suggested by previous studies. Finally, both skewness and kurtosis appear to be major sources of nonnormality.

Next we turn our attention to the market model residuals which are the

[9] The only notable difference is Bethlehem Steel. This difference is possibly attributable to corrections and updates made in each new version of the CRSP tapes.

Table I

A Comparison of Fama's Studentized Range Test of Normality on the Dow Jones 30 Industrials with Alternative Tests of Normality (Monthly Returns for January 1951–June 1968; $T = 210$)

SR is the Studentized range, W' the Shapiro-Francia statistic, $\sqrt{b_1}$ a measure of skewness, b_2 a measure of kurtosis, and $\bar{R}$ the average return. The Studentized range and average return subscripted with the letter F (column 1 and column 6) are taken from Fama (1976, p. 34). Note that the firms are listed under the company name in effect at the time of Fama's study.

Name	(1) SR_F	(2) SR	(3) W'	(4) $\sqrt{b_1}$	(5) b_2	(6) $\bar{R}_F$	(7) $\bar{R}$
Allied Chemical	7.75***	7.748***	0.967***	0.675***	5.270***	0.61%	0.61%
Alcoa[a]	7.00***	6.997***	0.991	0.024	3.396*	1.13	1.13
American Can	5.86	5.871	0.991	0.282**	3.200	0.84	0.85
AT & T	7.25***	7.253***	0.939***	0.650***	6.056***	0.81	0.82
American Tobacco	5.65	5.649	0.993	0.196	2.964	0.97	0.98
Anaconda	5.18	5.180	0.996	0.162	2.799	1.20	1.21
Bethlehem Steel	7.30***	5.595	0.981***	0.531***	3.530*	1.27	1.11
Chrysler	6.51**	6.516**	0.989*	0.261*	3.308	1.31	1.31
Du Pont	5.87	5.867	0.987**	0.386***	3.556*	0.91	0.92
Eastman Kodak	6.49**	6.534**	0.988*	0.292**	3.740**	1.75	1.75
General Electric	6.35	6.353	0.989*	0.305**	3.561*	1.23	1.24
General Foods	7.47***	7.473***	0.976***	0.454***	4.681***	1.40	1.40
General Motors	7.00***	6.920***	0.973***	0.670***	4.279***	1.39	1.40
Goodyear	5.62	5.622	0.991	0.318**	3.347	1.69	1.69
International Harvester	6.03	6.030	0.990	0.240*	3.047	0.88	0.89
International Nickel	6.91***	6.909***	0.992	0.071	3.488*	1.33	1.33
International Paper	5.77	5.777	0.993	0.247*	3.232	0.97	0.98
Johns Manville	5.68	5.678	0.980***	0.530***	3.799**	1.01	1.01
Owens Illinois	5.87	5.870	0.992	0.115	2.985	0.96	0.97
Procter & Gamble	6.12	6.124	0.992	0.085	3.325	1.17	1.18
Sears	6.49**	6.492**	0.979***	−0.130	4.120***	1.39	1.39
Standard Oil (CA)	5.71	5.715	0.995	0.161	2.884	1.06	1.06
Standard Oil (NJ)	5.73	5.730	0.982***	0.500***	3.251	1.21	1.22
Swift & Co.	5.99	5.995	0.991	0.322**	3.423*	0.76	0.76
Texaco	5.25	5.249	0.997	−0.015	2.555	1.48	1.49
Union Carbide	4.82	4.824	0.990	0.245*	2.582	0.61	0.61
United Aircraft	6.51**	6.517**	0.986**	0.366***	3.838**	1.59	1.59
U.S. Steel	7.61***	7.614***	0.965***	0.631***	5.315***	0.92	0.92
Westinghouse	5.13	5.131	0.989*	0.218*	2.714	1.16	1.17
Woolworth	6.88***	6.876***	0.943***	0.835***	5.559***	0.81	0.82

[a] Alcoa began trading on the NYSE on June 1951; thus, $T = 204$ in this case.

*, **, *** Significant at the 0.10, 0.05, and 0.01 levels, respectively.

variables of interest in the GRS test. We begin by examining the distribution of estimated monthly market model residuals for individual securities. To try to minimize the problem of nonstationarity in the market model parameters, the time period January 1931 to December 1985 is divided into eleven nonoverlapping sixty-month periods. Within each period, equation (1) is estimated using ordinary least squares regression for each security listed on the CRSP tape (with complete data). Then, the sixty estimates of market model residuals are used to calculate

the SR, W', $\sqrt{b_1}$, and b_2 statistics and to test the null hypothesis of normality. The results are summarized in Table II, which indicates the percentage of individual securities within each five-year period for which the normality hypothesis is rejected (at the five percent level). Overall, the results indicate that significant nonnormalities do exist in the estimated residuals of the market model when using individual securities. In particular, the following points from Table II are worth highlighting:

(i) In every period, and for all four test statistics, far more securities indicate significant nonnormality than the five percent that would be expected under the null hypothesis.

(ii) The percentage of securities displaying significant nonnormality differs from period to period. In particular, our results confirm those of Fama (1976) that the pre-World War II period exhibited greater nonnormal behavior than the postwar period. (See especially the 1936–1940 period.) Nevertheless, the results indicate that even in the postwar period at least one third of all securities examined exhibit significant departures from normality in the estimated residuals. For some periods (e.g., 1976–1980), more than half the securities display significant nonnormality.

(iii) The results of Table II confirm those of Table I in indicating that W', $\sqrt{b_1}$, and b_2 tests produce a greater number of rejections than the SR when testing for normality using stock return data. In particular, W' and b_2 both result in a greater number of rejections than the SR in every one of the eleven periods examined, while $\sqrt{b_1}$ demonstrates more rejections than the SR in nine of the eleven periods.

Table II

Distributional Tests for Market Model Residuals of Individual Securities by Period

Percentage of cases where the null hypothesis of a normal distribution is rejected at the $\alpha = 0.05$ level. Tests are based on the estimated residuals from equation (1) using sixty monthly observations. N is the number of individual securities tested in each period, SR the studentized range, W' the Shapiro-Francia statistic, $\sqrt{b_1}$ a measure of skewness, and b_2 a measure of kurtosis. All securities appearing on the CRSP monthly files with complete data for each period are included in the tests.

Period	N	SR	W'	$\sqrt{b_1}$	b_2
1/31–12/35	322	46.9%	54.7%	32.9%	57.5%
1/36–12/40	540	55.0	85.7	84.6	83.1
1/41–12/45	603	51.2	59.7	49.1	64.8
1/46–12/50	728	32.8	48.6	48.5	44.4
1/51–12/55	781	19.2	26.1	31.5	25.7
1/56–12/60	847	23.5	41.6	50.1	34.5
1/61–12/65	812	20.3	39.8	44.2	31.7
1/66–12/70	763	24.8	31.2	35.0	34.3
1/71–12/75	878	13.7	29.5	38.3	22.6
1/76–12/80	979	42.8	56.7	55.4	52.6
1/81–12/85	951	33.1	36.1	33.3	37.6
Average		33.0	46.6	45.7	44.2

(iv) The directional tests, when combined, are generally consistent with the Shapiro-Francia W' test (in terms of the number of rejections). In addition, the directional tests $\sqrt{b_1}$ and b_2 provide a characterization of the nonnormality in the data. Consequently, when examining the distribution of portfolio residuals in subsequent sections, results are reported only for the $\sqrt{b_1}$ and b_2 statistics.

Portfolios of assets rather than individual securities are frequently used when employing the GRS test. Hence, the final tests of normality in this section are performed on the market model residuals for portfolios rather than individual securities. For each nonoverlapping time period, twenty beta-sorted portfolios were formed (i.e., the securities within each period were ranked on beta, and the first portfolio contained the lowest beta securities, while the 20th portfolio included the highest beta securities).[10] For each portfolio the market model parameters were estimated and the estimated residuals computed using equation (1). The directional test statistics $\sqrt{b_1}$ and b_2 were then computed for each portfolio within each period, and the results are summarized in Table III. Due to the vast amount of data, results are reported only for four periods: the three periods indicating most nonnormality in Table II (1936–1940, 1941–1945, and 1976–1980) and the most recent period where nonnormalities are relatively low (1981–1985). In addition, Table III provides results for the same four time periods using size-sorted rather than beta-sorted portfolios (i.e., portfolio 1 includes the smallest capitalization firms within each period, and portfolio 20 includes the largest capitalization firms).

Several interesting results are apparent from Table III.

(i) Consistent with previous studies, the portfolio residuals indicate fewer nonnormalities in percentage terms than do the individual securities. This is true for all four periods analyzed and for both the $\sqrt{b_1}$ and b_2 tests.

(ii) Nevertheless, even for the period selected with minimal nonnormalities, for the size-sorted portfolios the extent of nonnormality is still significantly more than expected under the null hypothesis.

(iii) The residuals of the size-sorted portfolios generally indicate more nonnormality than do the residuals of the beta-sorted portfolios. Thus, our results indicate that the portfolio formation criterion may affect the degree of nonnormality in the residuals. However, significance within the size-sorted groupings appeared evenly spread across the twenty portfolios, and, thus, there is no evidence that departures from normality are more prevalent in portfolios of smaller or larger companies. Similar comments apply to the beta-sorted portfolios.

(iv) In almost every period examined in Table III, both significant skewness and kurtosis are evident for some of the portfolios.

Overall, the results presented in this section confirm that market model residuals can display significant nonnormality for both individual securities and

[10] For cases where the total sample of securities (M) is not evenly divisible by twenty, the first k portfolios contain one additional security, with $k = M - INT(M/20) \times 20$.

Table III

Residual Skewness ($\sqrt{b_1}$) and Kurtosis (b_2) by Test Period for Beta-Sorted and Size-Sorted Portfolios

Statistics for each portfolio are calculated using the sixty monthly market model residuals from equation (1). Portfolios are based on all securities with complete data for a given time interval and are reported in ascending order. $\sqrt{b_1}$ is a measure of skewness, and b_2 is a measure of kurtosis.

		1936–1940		1941–1945		1976–1980		1981–1985	
Portfolio		$\sqrt{b_1}$	b_2	$\sqrt{b_1}$	b_2	$\sqrt{b_1}$	b_2	$\sqrt{b_1}$	b_2
Beta-Sorted	1	−1.85*	9.13*	0.59*	5.68*	0.40	3.62	0.27	3.10
	2	−2.51*	12.03*	0.20	4.61*	0.36	3.92	−0.26	2.70
	3	−1.36*	6.62*	−0.18	4.09*	0.54*	5.35*	−0.05	2.50
	4	−1.02*	4.00*	0.18	3.42	−0.40	4.59*	−0.38	2.95
	5	−0.75*	3.73	−0.02	3.07	−0.46	4.65*	−0.28	2.75
	6	−1.84*	11.60*	0.23	3.93*	0.59*	4.60*	0.01	3.12
	7	−0.07	2.76	0.19	3.37	0.42	3.41	−0.12	2.45
	8	−0.22	2.86	−0.27	3.52	−0.10	2.69	−0.37	3.24
	9	−0.87*	4.37*	−0.06	3.39	−0.12	2.18	0.10	3.30
	10	−1.11*	8.64*	−0.15	2.79	0.25	3.81	0.02	3.68
	11	−1.12*	6.74*	0.61*	2.82	0.35	2.75	−0.39	3.44
	12	0.40	3.17	0.30	3.79	−0.59*	2.90	0.42	3.04
	13	−0.23	2.54	1.01*	5.83*	−0.24	2.82	−0.68*	3.03
	14	0.88*	5.49*	0.79*	4.38*	0.24	3.23	0.09	3.81
	15	0.44	2.48	−0.16	2.65	0.50*	4.42*	0.05	2.69
	16	1.36*	10.12*	0.44	3.86	0.38	3.64	0.13	2.72
	17	0.43	4.00*	−0.18	2.39	0.09	3.13	0.24	2.44
	18	2.79*	16.46*	0.82*	6.63*	0.02	3.25	−0.18	2.26
	19	0.16	3.07	0.16	3.64	−0.12	2.93	0.18	2.93
	20	0.32	2.33	0.13	4.08*	0.34	2.74	0.21	2.59
Average[a]		0.99	6.11	0.33	3.90	0.33	3.53	0.22	2.94
Size-Sorted	1	2.20*	12.49*	0.83*	4.57*	1.21*	6.16*	−0.11	3.13
	2	0.74*	3.04	0.18	3.04	0.17	2.85	0.86*	5.62*
	3	0.24	2.92	−0.19	2.60	−0.17	3.44	0.22	3.36
	4	1.94*	11.88*	0.40	3.10	0.47	3.57	0.47	2.92
	5	0.73*	3.93*	−0.14	2.43	0.89*	4.35*	0.72*	4.20*
	6	0.50*	3.07	0.30	2.86	−0.54*	4.73*	−0.39	3.05
	7	1.00*	6.41*	0.20	3.73	0.33	3.50	0.53*	5.16*
	8	−0.22	4.02*	−0.48	4.11*	0.75*	3.89	−0.09	3.44
	9	−0.64*	4.65*	−1.70*	8.80*	0.20	2.98	0.05	2.84
	10	0.26	3.25	−0.46	4.98*	0.52*	3.70	−0.10	3.07
	11	−0.16	3.54	−0.41	3.65	0.19	2.98	−0.85*	5.00*
	12	−0.79*	4.30*	−0.39	3.21	0.14	2.75	0.65*	3.93*
	13	−0.22	4.37*	−0.17	4.25*	−0.11	3.44	0.26	2.35
	14	−1.84*	9.31*	−1.97*	10.99*	0.06	2.75	0.21	2.84
	15	−1.09*	4.90*	−0.37	3.27	−0.05	3.27	0.07	2.68
	16	−1.21*	6.55*	−0.09	4.31*	−0.63*	3.57	0.03	3.20
	17	0.18	4.18*	0.07	5.10*	0.24	3.12	0.04	2.48
	18	−0.26	2.78	−0.06	2.89	−0.68*	4.49*	−0.48	3.63
	19	−0.41	3.45	−0.67*	4.54*	−0.53*	5.07*	−0.37	3.06
	20	−1.13*	5.75*	−0.66*	5.90*	0.04	4.98*	−0.24	3.07
Average[a]		0.79	5.24	0.49	4.42	0.40	3.78	0.34	3.45

[a] Average across the twenty portfolios. For $\sqrt{b_1}$, the average of the absolute value is reported.

* Significant at the 0.05 level. Critical values: $|\sqrt{b_1}| = 0.49$, $b_2 = 3.93$.

portfolios. Consequently, the effect of such nonnormalities on the significance level and power of parametric tests based on market model residuals should be an important concern of empirically based research. In the following section, this issue is examined in the context of the GRS test of efficiency of a given portfolio.

B. Robustness of the GRS Test

The results of the simulation are summarized in Table IV in the columns labeled "Nonnormal." In addition, Table IV lists the theoretical significance level and power assuming normality of the underlying residuals (columns labeled "Theoretical") as well as simulation results where the errors were generated from a multivariate normal distribution with the specified covariance structure (columns labeled "Normal"). The latter two serve as a comparison for the simulation

Table IV

Rejection Frequencies for Simulations for the GRS Test

The percentage of rejections of the null hypothesis that $\alpha_{ip} = 0$ for all i using the GRS test is reported. Empirical rejection frequencies are based on 5,000 replications. The theoretical rejection rate under the null and the alternative (columns 1 and 4) is calculated using the noncentral F, with the noncentrality parameter estimated as in equation (12) of MacKinlay (1987). Columns 2 and 5 are rejection rates based on simulations where the market model residuals are generated from a multivariate normal distribution. Columns 3 and 6 are the rejection rates from the simulations where the market model residuals are generated with nonnormalities replicating the observed sample levels of skewness and kurtosis. Column 7 is the same as column 6, with the exception that the critical value for the test is now based on the empirically determined five percent value (using the estimates from the simulation of column 3). For columns 2, 3, 5, and 6, the critical value for the test is given by a central $F_{20,39}$. The first three periods (1/36–12/40, 1/41–12/45, 1/76–12/80) represent periods with relatively high levels of observed nonnormalities. The 1/81–12/85 period is reported as a case where the nonnormalities were relatively low. (See Table III.)

Period	Null ($\gamma = 0$)			Alternative ($\gamma = 0.10$)			
	(1) Theoretical	(2) Normal	(3) Nonnormal	(4) Theoretical	(5) Normal	(6) Nonnormal	(7) Nonnormal—(Empirical Distribution)
Beta-Sorted Portfolios							
1/36–12/40[a]	5.00%	5.20%	10.76%**	11.51%	11.14%	27.80%**	17.36**
1/41–12/45	5.00	5.14	5.30	16.75	16.94	19.48**	18.66
1/76–12/80	5.00	5.26	5.56*	20.57	20.16	22.10**	20.18
1/81–12/85	5.00	5.32	5.26	20.93	21.00	22.32**	21.84
Size-Sorted Portfolios							
1/36–12/40	5.00%	4.82%	8.50%**	10.21%	10.48%	20.10%**	13.54**
1/41–12/45[b]	5.00	4.90	5.74**	17.41	17.62	22.38**	20.44**
1/76–12/80	5.00	5.08	6.02**	10.61	10.72	12.22**	10.58
1/81–12/85	5.00	4.92	5.44	13.80	13.22	13.50	12.62

[a] The simulated levels of skewness and kurtosis were reduced for portfolios 2 and 6 in period 1/36–12/40 for the beta-sorted sample by twenty percent (from the levels reported in Table III) to facilitate the simulation method. (See footnotes 4 and 7.)

[b] The simulated levels of skewness and kurtosis were reduced for portfolio 14 in period 1/41–12/45 for the size-sorted sample by ten percent (from the levels reported in Table III) to facilitate the simulation method.

*,** Significantly different from the theoretical rate at the 0.10 and 0.05 levels, respectively.

results and as a validation check on the simulation process. The final column of Table IV (labeled "Nonnormal—(Empirical Distribution)") presents power results where the critical value is determined as the 95th percentile of the simulated distribution under the null (i.e., the size of the test is maintained at or close to five percent).

An analysis of the results in Table IV indicates consistency of the simulation process, with multivariate normal errors providing both significance level ($\gamma = 0$) and power ($\gamma = 0.10$) values close to the theoretical levels (comparing column 1 with 2 and column 4 with 5). For the nonnormal errors, the significance level varies between 5.26 percent and 10.76 percent (column 3) and is significantly different from the specified level of five percent in five of the six cases in which substantial nonnormalities were apparent in the underlying data. Although the misstatement of the size of the test was statistically significant in these five of six cases, only one period (1936–1940) had deviations that could reasonably be considered problematic. Not surprisingly, this is the period with most deviations from normality. Moreover, the rejection frequency was not significantly different from the specified five percent level in both cases involving the most recent period (1981–1985) in which only small deviations from normality were evident.

The tests on the significance level of the GRS test indicate a higher rejection frequency than the expected level of five percent. Thus, it is not surprising that the corresponding simulation results for the nominal power (column 6) reveal that nonnormalities increase the empirical power of the test—in every case examined the observed power is greater than the theoretical power (calculated under the assumption of normality), significantly so in seven of the eight cases examined. As previously mentioned, the tests from this column cannot determine whether this is caused by an actual increase in power or by the misspecification of the size of the test.

A more appropriate examination of the effect of nonnormalities on the power of the GRS test is provided by setting the critical value equal to the 95th percentile of the empirical null distribution. These results (column 7) clearly indicate that much of the apparent increase in power evident in column 6 is caused by the misspecification of the size of the test. Nevertheless, in three cases the power remains significantly greater than the theoretical power calculated under the assumption of normality.

Thus, in general, our results indicate that the actual significance level and power of the GRS test are not markedly affected by reasonable levels of nonnormality. However, for more extreme deviations from normality observable in the data, this is not true and both the significance level and the power can be seriously misstated. Whether such extreme deviations are likely to occur in market model residuals will be discussed in the following section.

C. Implications for Tests Using Random Samples of Market Data

While the above results show that the GRS test is sensitive to certain large deviations from normality, the question remains as to whether these sample extremes are representative of the population values. In this regard it is important

to note that our results overstate the sensitivity of the test since our simulations are not based on a random choice of five-year intervals but include the three intervals that exhibit the greatest ex post skewness and kurtosis. Given that the choice of interval is based on ex post measures, it follows that these intervals might correspond to periods in which the population skewness and kurtosis are overestimated by the sample estimates. Thus, the true (population) skewness and kurtosis are likely to be less than the parameters specified in the simulation process, especially in the 1936–1940 period.

To provide some insight into the distribution of skewness and kurtosis present in five-year market model residuals, the following procedure was adopted.

(i) A random starting date was selected from the period 1/31 to 1/81.
(ii) As in the previous samples, twenty beta-sorted portfolios were formed using all securities with data available for the ten-year interval required for the estimation and sample periods. (As before, the process was also repeated for size portfolios.)
(iii) For each of the twenty portfolios, market model parameters were estimated for the five-year sample period and the sample skewness and kurtosis of the estimated residuals computed.
(iv) The average absolute skewness and the average kurtosis across the twenty portfolios were computed.[11]
(v) Steps (i) to (iv) were repeated 1,000 times to obtain empirical distributions for the average absolute skewness and average kurtosis over the period 1931–1985.

Some summary statistics for the distribution of average absolute skewness and average kurtosis are presented in Table V. It is important to note that this procedure provides only an approximation to the distributions of these measures for two reasons. First, it is based on only 1,000 replications. Second, because of data limitations, these 1,000 replications result from overlapping periods (and obviously stocks that are repeatedly selected), and, hence, they are not independent. Nevertheless, the empirical distributions do provide some indication of the levels of absolute skewness and kurtosis that can be expected in actual stock data.

The results in Table V indicate that the mean levels of absolute skewness and kurtosis over random five-year periods are fairly close to those used in our simulations for the periods 1941–1945 and 1976–1980. Thus, the results corresponding to these periods probably provide the best indication of the likely effects of nonnormalities in actual samples on the GRS test. Recalling the results from Table IV, this implies that there is probably some misstatement of the size of the test. However, this is likely to be small (i.e., given a specified size of five percent the actual size will be between five and six percent). As regards the power of the GRS test, the results in Table IV imply that the nonnormalities inherent in random samples might increase the power. Once again, however, such an

[11] Since we are concerned with skewness deviations from the normal value of zero, we examine the distribution of absolute skewness. This also facilitates comparison with the average values reported in Table III.

Table V

Summary Statistics for Empirical Distributions of Average Absolute Skewness and Average Kurtosis

For the period 1931–1985, 1,000 random sample periods were selected. For each of the 1,000 sample periods, twenty portfolios were formed using all available securities, and the average value of absolute skewness and kurtosis for the market model residuals across the twenty portfolios was calculated. The distribution statistics reported are based on these 1,000 average values (determined separately for the two different sorting methods).

	Average Absolute Skewness		Average Kurtosis	
Parameter	Beta-Sorted	Size-Sorted	Beta-Sorted	Size-Sorted
Mean	0.459	0.459	4.062	4.188
SD	0.360	0.367	1.478	1.598
Minimum	0.069	0.071	2.349	2.368
$P_{0.05}$[a]	0.130	0.130	2.533	2.540
$P_{0.10}$	0.150	0.150	2.598	2.626
$P_{0.50}$	0.305	0.335	3.459	3.677
$P_{0.90}$	0.923	0.922	5.968	6.220
$P_{0.95}$	1.291	1.185	6.950	7.423
Maximum	1.989	1.984	10.344	10.146

[a] Percentiles. $P_{0.05}$ refers to the lowest five percent of observations.

increase is likely to be small. Thus, we conclude that the levels of nonnormality typically present in stock market data are unlikely to affect materially either the size or the power of the GRS test.

D. Simulation Biases

It is important to note that our simulation results are potentially biased against finding misspecification. This is attributable to the sampling moments of moments problem. Under normality, simulated skewness and kurtosis are asymptotically unbiased estimators of the targeted level of skewness and kurtosis, respectively. This is not true, however, in the nonnormal case. As noted by Bowman and Shenton (1986), relatively little is known in general about the distributions of $\sqrt{b_1}$ and b_2 in the nonnormal case. Our simulations indicated that, for nonnormal data, the average values of the sampling estimators of skewness and kurtosis were always closer to normality than the targeted values (i.e., skewness closer to zero and kurtosis closer to three). It is possible using simulation methods to provide an ad hoc adjustment to the targeted values of skewness and kurtosis in the simulations so that the sample values are equal to the targeted values. We did experiment with this adjustment on the 1981–1985 data—the rejection frequencies increased from 5.26 to 5.38 under the null and from 22.32 to 22.66 under the alternative. Because of the ad hoc nature of the

adjustment and the small effect on the results, we have not incorporated the procedure in Table IV.[12]

V. Conclusions

Many studies have relied on the stylized fact that the distribution of monthly market model residuals is "reasonably normal." Whether the higher levels of nonnormalities found in this study using more powerful tests violate the bounds of "reasonable" is not the critical point of this study. More important is the result that existing nonnormalities generally do not have a substantive effect on the power and size of multivariate tests of asset pricing.

While the GRS test is not sensitive to minor deviations from normality, the effect of higher levels of nonnormality can be sizable. Although the most striking misspecifications are associated with early periods when the market was highly volatile, recent experience in the market suggests that these cases are not necessarily isolated artifacts of history. A general rule of thumb, based on the average values of skewness and kurtosis across the samples (as reported in Table III) and the associated results in Table IV, is that, when the residual diagnostics indicate average values for the sample that are within the range of the critical values for $\sqrt{b_1}$ and b_2 (0.49 and 3.93, respectively, for $T = 60$), then the size/power characteristics of the test are reasonably robust.

This study only considers the multivariate test of GRS. Any alternative methodological tests of asset pricing should similarly address the size and power issue. Additionally, in any of these applications, conclusions based on traditional hypotheses tests must be carefully considered in light of distributional diagnostics.

[12] Additionally, Huang and Bolch (1974) show that skewness and kurtosis of OLS residuals will never exceed and generally be less than the skewness and kurtosis of the true error term, a property referred to as supernormality. Consequently, our results might understate the effects of observed nonnormalities on the GRS test since our simulated nonnormalities mimic but tend to understate the actual level of skewness and kurtosis in the original sample. However, using an adjustment proposed by Huang and Bolch, it was found that the effect on our measures of skewness and kurtosis was relatively trivial.

REFERENCES

Anderson, Theodore W., 1984, *An Introduction to Multivariate Statistical Analysis* (John Wiley & Sons, Inc., New York).

Arrow, Kenneth J., 1960, Decision theory and the choice of a level of significance for the t-test, in Ingram Olkin et al., eds: *Contributions to Probability and Statistics* (Stanford University Press, Stanford, CA).

Bachelier, L. J. B. A., 1964, Theory de la speculation, in Paul Cootner, ed.: *The Random Character of Stock Market Prices* (M.I.T. Press, Cambridge, MA).

Blume, Marshall E., 1968, The assessment of portfolio performance: An application of portfolio theory, Ph.D. dissertation, University of Chicago.

Bookstaber, Richard M. and James B. McDonald, 1987, A general distribution for describing security price returns, *Journal of Business* 60, 401–424.

Bowman, K. O. and L. R. Shenton, 1975, Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2, *Biometrika* 62, 243–250.

——— and L. R. Shenton, 1986, Moment ($\sqrt{b_1}$ and b_2) techniques, in Ralph B. D'Agostino and Michael A. Stephens, eds.: *Goodness-of-Fit Techniques* (M. Dekker, New York).

Burgstahler, David, 1987, Inference from empirical research, *Accounting Review* 62, 203–214.

Clark, Peter K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.

David, H. A., H. O. Hartley, and E. S. Pearson, 1954, The distribution of the ratio, in a single normal sample, of range to standard deviation, *Biometrika* 41, 482–493.

D'Agostino, Ralph B. and E. S. Pearson, 1973, Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika* 60, 613–622.

Fama, Eugene F., 1976, *Foundations of Finance* (Basic Books, Inc., New York).

——— and Richard Roll, 1971, Parameter estimates for symmetric stable distributions, *Journal of the American Statistical Association* 66, 331–338.

——— and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Fleishman, Allen I., 1978, A method for simulating non-normal distributions, *Psychometrika* 43, 521–532.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1986, A test of the efficiency of a given portfolio, research paper #853, Graduate School of Business, Stanford University.

——— and Jay Shanken, 1987, Subperiod aggregation and the power of multivariate tests of portfolio efficiency, *Journal of Financial Economics* 19, 389–394.

Gnanadesikan, R., 1977, *Methods for Statistical Data Analysis of Multivariate Observations* (John Wiley & Sons, Inc., New York).

Huang, Cliff J. and Ben W. Bolch, 1974, On the testing of regression disturbances for normality, *Journal of the American Statistical Association* 69, 330–335.

Kandel, Shmuel and Robert F. Stambaugh, 1987, On correlations and inferences about mean-variance efficiency, *Journal of Financial Economics* 18, 61–90.

Kendall, Maurice G. and Alan Stuart, 1967, *The Advanced Theory of Statistics Vol. 2* (Hafner Publishing Co., New York).

Lehmann, Erich L., 1986, *Testing Statistical Hypotheses* (John Wiley & Sons, Inc., New York).

MacKinlay, A. Craig, 1987, On multivariate tests of the CAPM, *Journal of Financial Economics* 18, 341–371.

Mandelbrot, Benoit, 1963, The variation of certain speculative prices, *Journal of Business* 36, 394–419.

Mardia, K. V., 1980, Tests of univariate and multivariate normality, in P. R. Krishnaiah, ed.: *Handbook of Statistics I: Analysis of Variance* (North Holland Publishing Company, Amsterdam, North Holland).

Pearson, E. S., Ralph B. D'Agostino, and K. O. Bowman, 1977, Tests for departure from normality: Comparison of powers, *Biometrika* 64, 231–246.

Shanken, Jay, 1985, Multivariate tests of the zero-beta CAPM, *Journal of Financial Economics* 14, 327–348.

———, 1987a, Multivariate proxies and asset pricing relations: Living with the Roll critique, *Journal of Financial Economics* 18, 91–110.

———, 1987b, A Bayesian approach to testing portfolio efficiency, *Journal of Financial Economics* 19, 195–215.

Shapiro, S. S. and M. B. Wilk, 1965, An analysis of variance test for normality (complete samples), *Biometrika* 52, 591–611.

———, M. B. Wilk, and H. J. Chen, 1968, A comparative study of various tests for normality, *Journal of the American Statistical Association* 63, 1343–1372.

——— and R. S. Francia, 1972, An approximate analysis of variance test for normality, *Journal of the American Statistical Association* 67, 215–216.

Tadikamalla, Pandu R., 1980, On simulating non-normal distributions, *Psychometrika* 45, 273–279.

Vale, C. David and Vincent A. Maurelli, 1983, Simulating multivariate nonnormal distributions, *Psychometrika* 48, 465–471.