A Test for Multivariate Normality in Stock Returns

Matthew Richardson, Tom Smith

Matthew Richardson
University of Pennsylvania

Tom Smith
Duke University

A Test for Multivariate Normality in Stock Returns*

I. Introduction

A standard assumption in theoretical and empirical research in finance is that relevant variables (e.g., stock returns) have multivariate normal distributions. For example, in tests of mean-variance efficiency, small sample results have been derived under this assumption (see Mac-Kinlay 1987; Gibbons, Ross, and Shanken 1989). Moreover, justification for a number of asset pricing models has its roots in the multivariate normal assumption. Perhaps not surprisingly then, there has been considerable focus on whether this assumption is appropriate. (Please see Fama [1965, 1976]; Blume [1968]; Officer [1971]; Clark [1973]; Harris [1986]; Bookstaber and McDonald [1987]; and Affleck-Graves and McDonald [1989] for examples of this literature.) With respect to stock returns, the conclusion generally has been that returns are not normally distributed, putting into doubt results that rely heavily on this assumption.

These conclusions, however, are based on univariate tests of normality. For example, Fama (1976) finds that the studentized range test re-

Previous research has investigated the multivariate normality of stock returns using tests based on the marginal distribution of returns. Due to the contemporaneous correlation across asset returns, these tests are difficult to interpret. We develop a general test procedure that takes account of the correlation across assets and that focuses on both the marginal and joint distributions of returns. We find highly significant evidence that stock returns and market-model residuals are nonnormal. Moreover, this nonnormality appears in both the marginal and joint distributions of asset returns.

jects the normal distribution for monthly returns for 14 of the 30 Dow Jones Industrials over the 1951–68 sample period. Since then, numerous other test statistics have been reported, providing similar rejections. Note that, if a random variable is not univariate normally distributed, then it cannot have come from a multivariate normal distribution. It would seem, therefore, that this test procedure and, perhaps even more important, the corresponding evidence against multivariate normality is valid. Fama (1976) points out, however, that since returns are contemporaneously correlated the statistics will not be independent. The purpose of this article is to provide a procedure for calculating multivariate test statistics that takes account of this cross-sectional dependence.

Specifically, we develop a procedure for testing whether a multivariate time series of observations has a multivariate normal (MVN) distribution. This procedure is based on Hansen's generalized method of moments (GMM) approach. Intuitively, the MVN distribution imposes restrictions on the marginal and joint moments of the multivariate time series in terms of a relatively small number of parameters: the means, variances, and cross correlations. Overidentifying restrictions can be formed to test whether these restrictions hold for a given sample of observations. Interestingly enough, the cross-correlation parameters that need to be estimated are precisely the ones that Fama (1976) was worried about in reporting his results. Thus, the test statistics will be intuitively appealing as they incorporate this cross correlation directly.

The article is organized as follows: Section II motivates the analysis by studying the cross-dependence properties of individual statistics of particular interest to financial economists, namely, skewness and kurtosis. Section III develops a more general procedure for testing whether a multivariate series is MVN distributed. Section IV applies this procedure to test whether the residuals from market-model regressions are multivariate normally distributed. Section V discusses some extensions. Section VI concludes the article.

II.   On Tests for Univariate Normality

It is well known that, if a vector of asset returns R_t is MVN distributed, then each asset return R_{it} is univariate normally distributed. Therefore, although univariate normality does not imply multivariate normality, rejection of univariate normality is sufficient to reject the MVN condition. The mass of evidence suggesting that some individual stock returns do not come from univariate normal distributions would then seem to indicate stock returns are in fact not distributed as multivariate normals. With respect to tests for multivariate normality, however,

drawing inferences from univariate statistics can be misleading. The reason is that, given the correlation across assets, the univariate statistics will in general be correlated. This correlation suggests the need for a joint test across the asset returns being analyzed.

Consider two particular tests of normality, namely, the skewness and kurtosis measures:

$$\sqrt{T}S_i \equiv \sqrt{T}\frac{\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^3}{\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2\right]^{3/2}} \overset{asy}{\sim} N(0, 6)$$

and

$$\sqrt{T}K_i \equiv \sqrt{T}\left\{\frac{\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^4}{\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2\right]^2} - 3\right\} \overset{asy}{\sim} N(0, 24),$$

where R_{it} = return on asset i and $\hat{\mu}_i = 1/T\sum_{t=1}^{T} R_{it}$.

These statistical measures have been especially appealing to financial economists because they focus on properties of the distribution that are of low enough order to have an identifiable effect on asset returns and derivative securities. For example, Kraus and Litzenberger (1976) and Breeden (1986) investigate theoretical asset pricing models that employ third and fourth moments directly. Similarly, the mixture of distributions model (see, among others, Clark 1973; and Tauchen and Pitts 1983) implies excess kurtosis and skewness relative to the normal distribution. Finally, these measures have clear interpretations in terms of deviations from normality. That is, potential departures from the univariate normal null will point toward alternative distributions that do satisfy the skewness and kurtotic shapes, as was the case with Clark's (1973) mixture of distributions model and Fama's (1965) initial empirical investigation of stock prices.

In terms of empirical work, existing stylized facts from the literature are that continuously compounded returns are negatively skewed and leptokurtic. With respect to monthly returns, however, this evidence is considered especially weak; see, for example, Blume (1968), Officer (1971), Fama (1976), and, more recently, Affleck-Graves and McDonald (1989), among others. For example, in Fama's (1976) investigation, less than half of the Dow Jones firms have studentized ranges ex-

ceeding the 10% significance level. For motivational purposes, table 1 provides individual skewness and kurtosis tests for monthly returns of each Dow Jones 30 firm over the exact same sample period as Fama (1976).[1] Only 12 firms have statistically significant excess kurtosis at the 10% level, confirming conclusions reached in Fama (1976).[2] The skewness coefficients provide somewhat stronger evidence against normality. Over half of the firms display significant skewness coefficients.

Consider for the moment the appearance of excess kurtosis in some of the individual stock returns. There appear to be two possible explanations for this kurtosis. First, stock returns are actually drawn from some alternative distribution to the multivariate normal (perhaps a multivariate Student t or multivariate mixture of normals, both of which produce "excess" kurtosis). A second, more subtle, explanation is that the kurtosis patterns in stock returns may be spurious. The argument goes something like this: suppose stock returns are in fact MVN distributed. If we were to estimate the kurtosis of each stock return, then (by chance) we would expect some to exhibit excess kurtosis. If asset returns were cross-sectionally uncorrelated, then any excess kurtosis could be interpreted in terms of univariate statistics with mild adjustments. However, if asset returns are highly correlated (as they seem to be), then conditional on an asset exhibiting kurtosis we would expect, even under the null hypothesis of multivariate normality, other assets to also exhibit some degree of kurtosis. Thus, cross-sectional correlation across assets can lead to cross-sectional patterns of kurtosis in small samples. In interpreting the actual kurtosis results, therefore, the econometrician faces an identification problem: is the fact that kurtosis shows up in some assets due to spurious kurtosis coupled with the correlation pattern across asset returns or is it due to true kurtosis and the absence of normality in the returns' distributions?

However, if only a few stock returns exhibit kurtosis (as in table 1), it might suggest the MVN assumption is a good working approximation for stock returns. This type of reasoning can be misleading. Given the correlation across asset returns, it may be that estimates of excess kurtosis in only a few stocks provide substantial evidence against mul-

1. See the Appendix for a list of the Dow Jones 30 firms during the 1951–68 sample period.
2. Table 1 provides p-values (based on the asymptotic distribution) and Monte Carlo p-values. Monte Carlo p-values are based on a simulation where test statistics are calculated using 210 observations from a multivariate normal distribution with variance-covariance matrix equal to the empirical variance-covariance matrix of the 30 Dow Jones companies for the period January 1951 to June 1968. The Monte Carlo distribution is based on 5,000 repetitions. Tables 2–4 provide Monte Carlo p-values using the same simulation method.

TABLE 1   **Skewness and Kurtosis Statistics**

Company	Skewness	*p*-Value	Monte Carlo *p*	Kurtosis	*p*-Value	Monte Carlo *p*
ALD	.6745	.0001	.0000	2.2702	.0000	.0000
AA	.0255	.8818	.8820	.3964	.2478	.2104
AMB	.1994	.2381	.2334	−.0352	.9170	.9180
T	.6541	.0001	.0002	3.0597	.0000	.0000
A	.1640	.3318	.3272	−.2006	.5530	.5400
BS	.5334	.0016	.0024	.5323	.1154	.0882
CHV	.1644	.3308	.3256	−.1161	.7313	.7362
C	.2630	.1197	.1134	.3091	.3606	.3254
DD	.3895	.0212	.0236	.5576	.0991	.0766
EK	.2939	.0820	.0854	.7411	.0284	.0254
ESM	.3258	.0539	.0548	.4252	.2085	.1736
XON	.5021	.0030	.0042	.2525	.4552	.4314
GE	.3083	.0682	.0684	.5619	.0965	.0738
GF	.4571	.0069	.0074	1.6833	.0000	.0000
GM	.6733	.0001	.0000	1.2822	.0002	.0026
GT	.3216	.0571	.0588	.3488	.3022	.2636
N	.0755	.6553	.6584	.4883	.1486	.1136
IP	.2495	.1399	.1352	.2330	.4906	.4696
JM	.5320	.0016	.0024	.8008	.0179	.0184
NAV	.2441	.1486	.1464	.0479	.8874	.8878
OI	.1181	.4849	.4848	−.0145	.9659	.9680
PA	.2846	.0922	.0914	.2011	.5520	.5380
PG	.0881	.6022	.6048	.3250	.3364	.2986
S	−.1269	.4529	.4526	1.1189	.0009	.0058
TX	−.0108	.9491	.9476	−.4448	.1883	.1508
X	.6324	.0002	.0002	2.3180	.0000	.0000
UK	.2476	.1429	.1374	−.4170	.2173	.1846
UTX	.3674	.0297	.0308	.8382	.0132	.0156
WX	.2219	.1892	.1872	−.2845	.4000	.3700
Z	.8385	.0000	.0000	2.5626	.0000	.0000
χ^2_{30}	99.5404	.0000	.0000	286.7852	.0000	.0000

NOTE.—This table tests for normality of the Dow Jones 30 companies for the period January 1951–June 1968. Column 1 contains the ticker symbol of the corporation. Column 2 contains the skewness statistic, defined as the third central moment divided by $\hat{\sigma}^3$. Columns 3 and 4 contain this statistics table *p*-value and empirical *p*-value. Column 5 contains the kurtosis statistic, defined as the fourth central moment divided by $\hat{\sigma}^4$ minus 3. Columns 6 and 7 contain the table and Monte Carlo *p*-value of this statistic. The last row of this table contains the multivariate χ^2 test for normality across all companies. Monte Carlo *p*-values are based on a simulation where test statistics are calculated using 210 observations from a multivariate normal distribution with variance-covariance matrix equal to the empirical variance-covariance matrix of the 30 Dow Jones companies for the period January 1951–June 1968. Five thousand repetitions of the simulation are made.

tivariate normality. This is because the cross correlation across asset returns provides information about the accuracy and precision of each kurtosis estimate in the joint system.

It is possible to take into account the dependence between the univariate statistics when testing for normality. For example, consider the skewness and kurtosis measures for two assets *i* and *j*. Using results in Hansen (1982) and the procedure developed in the next section, the

following joint asymptotic distributions for the skewness and kurtosis measures can be derived:

$$\sqrt{T}\begin{pmatrix} S_i \\ S_j \\ K_i \\ K_j \end{pmatrix} \overset{asy}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 6 & 6\rho_{ij}^3 & 0 & 0 \\ 6\rho_{ij}^3 & 6 & 0 & 0 \\ 0 & 0 & 24 & 24\rho_{ij}^4 \\ 0 & 0 & 24\rho_{ij}^4 & 24 \end{pmatrix} \right),$$

where ρ_{ij} = the correlation between assets i and j.

Consider the sample correlation estimates between monthly asset returns of the Dow Jones 30 firms over the period 1951–68. These correlations vary from .0811 to .8425. Consider the two stocks with the highest correlation, namely, Bethlehem Steel (BS) and United States Steel (X). Using the joint asymptotic distribution of K_i and K_j, the correlation between the Bethlehem Steel's kurtosis measure and United States Steel's kurtosis is over 50%. Under the null hypothesis, this imposes sharp restrictions on the kurtosis patterns of these two stock returns. In general, conditional on one asset exhibiting "apparent kurtosis," under the null hypothesis that returns are normally distributed we would expect other correlated assets to exhibit similar kurtosis—in particular, the more closely correlated assets should have the most similar kurtosis. Thus, even though the magnitudes of kurtosis for each asset are important determinants of the distribution, an equally important factor is the pattern in kurtosis measures across assets.

One way to test for normally distributed returns in this environment is to form a joint test across asset returns. For example, let K be the N-vector of kurtosis measures for N assets, let $V(K)$ be the variance-covariance matrix of these kurtosis measures given above, and let A be an $M \times N$ matrix of constants. Then

$$T(AK)'[AV(K)A']^{-1}(AK) \overset{asy}{\sim} \chi_M^2. \tag{1}$$

One popular example of test restrictions for a joint test that $K_i = 0$ for all i is the Wald statistic. Specifically, let $A = I$, the $N \times N$ identity matrix. Then the test statistic in (1) reduces to

$$W \equiv TK'[V(K)]^{-1}K \overset{asy}{\sim} \chi_N^2. \tag{2}$$

Over the sample period 1951–68, the Wald test statistic in equation (2) is calculated for skewness and kurtosis restrictions across the 30 Dow Jones firms. These test results are provided in table 1. While the individual kurtosis tests imply normality may be a good approximation, the joint tests reject the multivariate normality of stock returns below the .0001 level of significance. The joint tests of skewness across stock

returns also strongly reject multivariate normality. In general, the joint tests suggest much less evidence of normally distributed monthly stock returns than do individual tests.

Note that the above statistics focus on the marginal distributions of individual asset returns. By correctly taking into account the correlation between these univariate statistics, rejection of marginal normality is sufficient to reject the MVN restriction. It should be pointed out, however, that for other applications detection of nonnormality through marginal normality tests may be difficult. Tests that exploit the multivariate structure should, however, be more sensitive to departures from the null. It is possible to incorporate this multivariate structure of asset returns directly by estimating implied cross moments of asset return distributions. The next section proposes a general procedure for testing whether a multivariate series conforms to an MVN distribution.

III. Multivariate Test: Theory

There are a number of existing procedures for testing whether multivariate series are MVN. For example, in relation to results in this article, Mardia (1970) proposes multivariate measures of skewness and kurtosis, which are special cases of MVN moment restrictions and, therefore, of the GMM procedure outlined below. Cox and Small (1978) test for whether two series (x_{it}, x_{jt}) are bivariate normal. They propose a Wald-type test on the t-statistics from regressions of x_{it} on x_{jt} and x_{jt}^2 (and vice versa). This procedure is similar in spirit to testing conditional moment restrictions, which also falls into the GMM methodology. In addition, multivariate generalizations of the popular Shapiro-Wilks and Kolmogorov-Smirnov tests have also been developed. All of these MVN test methodologies, however, have not yet been adopted toward applications in finance.[3] (See Mardia [1980] for an excellent survey of the literature on these and other tests for multivariate normality.)

This section develops an alternative procedure for testing whether a multivariate time series of observations has an MVN distribution. Let $\{R_t\}_{t=1}^T = \{R_{1t}, \ldots, R_{Nt}\}_{t=1}^T$ be an N-vector time series of observations from an independently and identically distributed (i.i.d.) multivariate distribution F.[4] If the multivariate series R_t conforms to F, then its moments (as long as they exist) should also conform to F's:

$$E[h(R_t, \theta)] = 0, \qquad (3)$$

3. An exception is Zhou (1991), who uses the multivariate skewness and kurtosis tests in order to examine multivariate normality of market-model regression residuals for industry portfolios.

4. The i.i.d. assumption can be weakened to simply stationarity and ergodicity. Since most previous test procedures for normality require the i.i.d. assumption, the weaker

where θ equals an M-vector of parameters governing F and $h(\cdot)$ is an R-vector of functional forms.

In large samples, under the null hypothesis that $R_t \overset{d}{\sim} F$, the sample moments of (3) will converge in mean square to zero:

$$g_T(\theta) \equiv \frac{1}{T} \sum_{t=1}^{T} h(R_t, \theta) \overset{T \to \infty}{\to} 0.$$

The idea behind the GMM procedure is to find the values of the unknown parameters θ that set the sample vector $g_T(\theta)$ equal to zero. This will not be possible if the system is overidentified, that is, if $M < R$. We can, however, set $M \times R$ linear combinations (denote A) of the R-vector $g_T(\theta)$ to zero:

$$Ag_T(\theta) = 0. \tag{4}$$

Hansen (1982) shows that the optimal choice of A in terms of minimizing the variance-covariance matrix of the parameter estimates $\hat{\theta}$ from (4) is $A = D_0' S_0^{-1}$, where $D_0 = E[\partial h(R_t, \theta)/\partial \theta]$ and $S_0 = E[h(R_t, \theta)h(R_t, \theta)']$. Of special interest to this article, Hansen also provides the following statistical results:

$$\sqrt{T}(\hat{\theta} - \theta) \overset{asy}{\sim} N(0, [D_0' S_0^{-1} D_0]^{-1}),$$

and

$$J_T \equiv Tg_T(\hat{\theta})' S_0^{-1} g_T(\hat{\theta}) \overset{asy}{\sim} \chi^2_{R-M}.$$

In practice, D_0 and S_0 are usually unknown; however, all that the theory requires are consistent estimators of D_0 and S_0 for the asymptotic normality and asymptotic χ^2 distribution results to hold. For example, one possible estimator for the asymptotic variance ($[D_0' S_0^{-1} D_0]^{-1}$) is ($[D_T' S_T^{-1} D_T]^{-1}$), where D_T and S_T denote the sample moment estimates.

Although the analysis in this article focuses on the MVN distribution, the procedure is applicable to any multivariate distribution as long as its moments (in fact, two times the highest-order moment looked at) exist and are finite.

---

assumptions are an attractive feature of the GMM test. The importance of relaxing the i.i.d. assumption can be seen in recent empirical work which suggests stock returns may be serially correlated (see, e.g., Lo and MacKinlay 1988) and heteroscedastic (see Schwert 1989). This is of particular importance given that normality can be maintained in the presence of serially correlated data. To coincide with the previous literature, however, we maintain the i.i.d. assumption throughout most of the article and relax it in Sec. VA.

A. Multivariate Normal Distribution

The MVN distribution expresses its moments in terms of relatively only a few parameters: the means, variances, and correlations between $R_{1t}, \ldots, R_{Nt}$. Therefore, many overidentifying (i.e., testable) restrictions can be placed on the data.

Without loss of generality, consider just two series, R_{it} and R_{jt}, which are bivariate normal. Under this assumption, the moment generating function is given by

$$M(t_i, t_j) = e^{t_i \mu_i + t_j \mu_j + 1/2(t_i^2 \sigma_i^2 + 2\rho_{ij} t_i t_j \sigma_i \sigma_j + t_j^2 \sigma_j^2)}.$$

We can obtain all the moments, $E[R_{it}^p R_{jt}^q]$ for all integers p and $q \geq 0$, by differentiating $M(t_i, t_j)$ p times with respect to t_i and q times with respect to t_j and then setting t_i and t_j equal to zero. Using this technique, it is possible to form more individual and joint moments than the five parameters (μ_i, μ_j, σ_i^2, ρ_{ij}, σ_j^2) needed for estimation. Therefore, using the procedure above, we can test whether (R_{it}, R_{jt}) are bivariate normal.

For example, consider the following sample moment conditions relating to the first four moments and corresponding cross moments of R_{it} and R_{jt} (note that sample moment conditions can be expressed for any higher-order moment—we focus on the third and fourth moments to coincide with the previous discussion on skewness and kurtosis):

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} R_{it} - \mu_i \\ R_{jt} - \mu_j \\ (R_{it} - \mu_i)^2 - \sigma_i^2 \\ (R_{jt} - \mu_j)^2 - \sigma_j^2 \\ (R_{it} - \mu_i)(R_{jt} - \mu_j) - \sigma_i \sigma_j \rho_{ij} \\ (R_{it} - \mu_i)^3 \\ (R_{jt} - \mu_j)^3 \\ (R_{it} - \mu_i)^2 (R_{jt} - \mu_j) \\ (R_{it} - \mu_i)(R_{jt} - \mu_j)^2 \\ (R_{it} - \mu_i)^4 - 3\sigma_i^4 \\ (R_{jt} - \mu_j)^4 - 3\sigma_j^4 \\ (R_{it} - \mu_i)^2 (R_{jt} - \mu_j)^2 - \sigma_i^2 \sigma_j^2 (1 + 2\rho_{ij}^2) \\ (R_{it} - \mu_i)^3 (R_{jt} - \mu_j) - 3\sigma_i^3 \sigma_j \rho_{ij} \\ (R_{it} - \mu_i)(R_{jt} - \mu_j)^3 - 3\sigma_i \sigma_j^3 \rho_{ij} \\ \vdots \end{pmatrix}, \quad (5)$$

where $\theta \equiv (\mu_i, \mu_j, \sigma_i^2, \rho_{ij}, \sigma_j^2)$.

With these restrictions alone, the econometrician has 14 moment conditions and only five parameters for estimation, leaving him with nine overidentifying restrictions to test. In addition, as assets are added, the number of testable restrictions increases by the rate $N(N - 1)/2$, where N is the number of assets. Using the GMM estimation procedure in equation (4), it is then straightforward to test these restrictions.

B.   Optimal GMM Estimators: Theory

For simplicity, consider the moment restrictions in equation (5) of Section IIIA. Under the null hypothesis that stock returns are MVN distributed, the derivative matrix and variance-covariance matrix D_0 and S_0 can be calculated analytically. In fact, they will have representations in terms of only the mean, variance, and correlation parameters θ. Using Hansen's results, it is possible to calculate the optimal GMM weights in equation (4), the 5×14 matrix $A^* = D_0' S_0^{-1}$.

The optimal weights given by A^* take on an especially interesting form. Specifically, consider partitioning the general A matrix into two matrices, one 5×5 and the other 5×9. In the optimal GMM case, it is possible to show that the 5×5 matrix is the identity matrix while the 5×9 matrix is a matrix of zeros—all the weight in estimation is placed on the first five moments. Using this result, it is possible to derive the optimal GMM estimators:

$$\hat{\mu}_i = \frac{1}{T} \sum_{t=1}^{T} R_{it},$$

$$\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=1}^{T} (R_{it} - \hat{\mu}_i)^2,$$

$$\hat{\mu}_j = \frac{1}{T} \sum_{t=1}^{T} R_{jt},$$

$$\hat{\sigma}_j^2 = \frac{1}{T} \sum_{t=1}^{T} (R_{jt} - \hat{\mu}_j)^2,$$

and

$$\hat{\rho}_{ij} = \frac{\frac{1}{T} \sum_{t=1}^{T} (R_{it} - \hat{\mu}_i)(R_{jt} - \hat{\mu}_j)}{\hat{\sigma}_i \hat{\sigma}_j}.$$

The intuition for this choice of weights (and hence the estimators)

is straightforward. Note that the GMM procedure chooses the weights that minimize the variance-covariance matrix for a given set of moment conditions. It is well known that maximum likelihood estimation (MLE) asymptotically achieves the Cramer-Rao lower bound. For the case of multivariate normality, the MLE estimates are simply the sample mean, variances, and correlations of the asset returns. These sample moments are completely described by the first five moments in equation (5). In fact, as long as the first five moments are included in estimation, the GMM procedure will always pick them out, irrespective of any other moment restrictions like those given in equation (5).

The econometrician can then substitute consistent estimates of θ (e.g., the sample estimates) into these expressions for S_0 to get the required consistent estimate, $S_0(\hat{\theta})$.[5] The $J_T(\hat{\theta})$ statistic then weights the nonzero moments in equation (5) (i.e., the higher-order moments) by this estimate of the variance-covariance matrix of the moment conditions, $S_0(\hat{\theta})$.

C. Cross-Moments Example: Results

To coincide with the analysis in Section II, we can identify the system of equations in (5) by adding skewness, kurtosis, and cross-moment parameters. Specifically, these identified restrictions in (5) imply the following cross-skewness and cross-kurtosis measures between assets i and j:

$$S_{ij} \equiv \frac{\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2(R_{jt} - \hat{\mu}_j)}{\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2\right]\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{jt} - \hat{\mu}_j)^2\right]^{1/2}},$$

and

$$K_{ij} \equiv \frac{\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2(R_{jt} - \hat{\mu}_j)^2}{\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{it} - \hat{\mu}_i)^2\right]\left[\dfrac{1}{T}\sum_{t=1}^{T}(R_{jt} - \hat{\mu}_j)^2\right]} - (1 + 2\hat{\rho}_{ij}^2),$$

5. While estimating S_0 this way will have no effect on the asymptotic distribution under the null, it will have different consequences for the small sample null distribution of the statistics as well as for the power of the statistics under alternative multivariate distributions. Section V C discusses this point in more detail.

where

$$\hat{\rho}_{ij} \equiv \frac{\displaystyle\sum_{t=1}^{T} (R_{it} - \hat{\mu}_i)(R_{jt} - \hat{\mu}_j)}{\left[\displaystyle\sum_{t=1}^{T} (R_{it} - \hat{\mu}_i)^2\right]^{1/2} \left[\displaystyle\sum_{t=1}^{T} (R_{jt} - \hat{\mu}_j)^2\right]^{1/2}}.$$

Using the result that the asymptotic variance of the parameter estimators is $[D_0' S_0^{-1} D_0]^{-1}$, the asymptotic distribution of the N-vector $S(i) = (S_{ij}, \ldots, S_{kl})$ and $K(i) = (K_{ij}, \ldots, K_{kl})$ can be derived. The typical elements of the variance-covariance matrix of the vector $S(i)$ are

$$\sqrt{T} \begin{pmatrix} S_{ij} \\ S_{kl} \end{pmatrix} \overset{asy}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4\rho_{ij}^2 + 2 & 2\rho_{ik}^2 \rho_{jl} + 4\rho_{ik}\rho_{il}\rho_{jk} \\ 2\rho_{ik}^2\rho_{jl} + 4\rho_{ik}\rho_{il}\rho_{jk} & 4\rho_{kl}^2 + 2 \end{pmatrix} \right)$$

The typical elements of the variance-covariance matrix of the vector $K(i)$ are

$$\sqrt{T} \begin{pmatrix} K_{ij} \\ K_{kl} \end{pmatrix} \overset{asy}{\sim} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \right.$$

$$\left. \begin{pmatrix} 4\rho_{ij}^4 + 16\rho_{ij}^2 + 4 & 4\rho_{ik}^2\rho_{jl}^2 + 16\rho_{ik}\rho_{il}\rho_{jk}\rho_{jl} + 4\rho_{il}^2\rho_{jk}^2 \\ 4\rho_{ik}^2\rho_{jl}^2 + 16\rho_{ik}\rho_{il}\rho_{jk}\rho_{jl} + 4\rho_{il}^2\rho_{jk}^2 & 4\rho_{kl}^4 + 16\rho_{kl}^2 + 4 \end{pmatrix} \right).$$

Note that the asymptotic covariance between $S(i)$ and $K(i)$ is zero. That is,

$$\text{cov}(S_{ij}, K_{kl}) = 0 \quad \forall\, i, j, k, \text{ and } l.$$

Similar to the kurtosis measures given in Section II, these measures will be highly correlated when the asset returns have high correlation. Consider two cross-kurtosis statistics, K_{ij} and K_{ik}. Suppose, for example, $\rho_{st} = .90$ for all s, t. In this case, over 80% of the variation in K_{ij} can be explained by K_{ik}. This imposes strong restrictions on cross-kurtosis measures in the data. We explore some of these restrictions below.

   Using the same data as Section II, we perform tests of multivariate normality that exploit the multivariate structure of asset returns. To coincide with the skewness and kurtosis measures estimated in Section II, we look at corresponding cross moments of skewness and kurtosis across the 30 Dow Jones firms. To keep down the number of restrictions, we choose one of the assets (e.g., Allied Corporation [ALD]) as a benchmark. This leads to 29 joint testable restrictions. Table 2 reports the individual measures $S_{1,j}$ and $K_{1,j}$ for $j = 2, \ldots, 30$ and the

TABLE 2   Coskewness and Cokurtosis Statistics

Company	Skewness	*p*-Value	Monte Carlo *p*	Kurtosis	*p*-Value	Monte Carlo *p*
AA	.2004	.0921	.0902	−.0015	.9938	.9926
AMB	.3977	.0002	.0006	.9129	.0000	.0004
T	.0786	.4875	.4896	−.0151	.9336	.9244
A	.2059	.0777	.0758	.1728	.3645	.3194
BS	.2809	.0239	.0254	.5105	.0158	.0146
CHV	−.1193	.2748	.2814	−.0454	.7896	.7784
C	.1229	.2582	.2648	.1831	.2782	.2334
DD	.1681	.1958	.1988	−.2096	.3549	.3092
EK	−.1005	.3797	.3798	.0341	.8531	.8458
ESM	.1485	.1680	.1678	.3134	.0591	.0448
XON	−.0897	.4151	.4130	.0486	.7778	.7658
GE	−.1457	.2103	.2132	−.2063	.2765	.2324
GF	−.0294	.7937	.7920	−.1295	.4682	.4312
GM	.2035	.0804	.0800	.4551	.0165	.0154
GT	.0263	.8316	.8302	−.0597	.7755	.7644
N	−.1191	.2783	.2846	−.0376	.8270	.8162
IP	.1355	.2736	.2802	.0984	.6389	.6142
JM	.3716	.0022	.0028	.8085	.0001	.0018
NAV	.1070	.3299	.3334	−.1121	.5143	.4816
OI	.0641	.5659	.5676	−.0134	.9395	.9300
PA	−.0354	.7436	.7436	−.1007	.5476	.5210
PG	.1464	.1679	.1676	.3100	.0556	.0410
S	.2233	.0487	.0510	.4568	.0117	.0128
TX	−.0804	.4653	.4644	−.0786	.6490	.6246
X	.2978	.0184	.0176	.7969	.0002	.0028
UK	.3158	.0137	.0140	.2229	.3144	.2698
UTX	.0788	.4498	.4452	−.1381	.3775	.3328
WX	.1667	.1674	.1674	.0101	.9601	.9522
Z	.2095	.0500	.0516	.4176	.0108	.0124
χ^2_{29}	98.7291	.0000	.0000	99.4022	.0000	.0036

NOTE.—This table tests for normality of the Dow Jones 30 companies for the period January 1951–June 1968. ALD is used as the reference asset. Column 1 contains the ticker symbol of the corporation. Column 2 contains the coskewness statistic. Columns 3 and 4 contain the table and Monte Carlo *p*-value of this statistic. Column 5 contains the cokurtosis statistic. Columns 6 and 7 contain the table and Monte Carlo *p*-value of this statistic. The last row of this table contains the multivariate χ^2 test for normality across all companies. Monte Carlo *p*-values are based on a simulation where test statistics are calculated using 210 observations from a multivariate normal distribution with variance-covariance matrix equal to the empirical variance-covariance matrix of the 30 Dow Jones companies for the period January 1951–June 1968. Five thousand repetitions of the simulation are made.

corresponding Wald statistics for joint tests of the hypothesis that $S_{1,j} = 0$ and $K_{1,j} = 0, j = 2, \ldots, 30$.

As with the individual kurtosis measures, the univariate cross-kurtosis measures provide weak evidence against normality, with nine of the 29 firms (i.e., ignoring the benchmark) displaying excess cross kurtosis at the 10% level. The cross-skewness measures provide similar results with only 10 of 29 firms displaying excess cross skewness at the 10% level. Similar to the joint tests given in table 1, the joint tests

across the cross-skewness and cross-kurtosis measures are significant below the .0001 level. Evidently, there is substantial evidence against multivariate normality at both the marginal and joint distributional levels.

IV. Empirical Example: Are Market-Model Residuals Multivariate Normal?

As Affleck-Graves and McDonald (1989) point out, the crucial assumption underlying many tests of asset-pricing theories is multivariate normality of market-model residuals. While some authors have found this assumption to be empirically unimportant (e.g., see MacKinlay 1985; and Affleck-Graves and McDonald 1989), others have shown that violation of this assumption can lead to incorrect inference (e.g., see MacKinlay and Richardson [1991] for a discussion of these different conclusions). Given the importance of this assumption, there have been surprisingly few empirical investigations of the distributional properties of the residuals.

Exceptions are Affleck-Graves and McDonald (1989) and Zhou (1991). Zhou (1991) uses multivariate skewness and kurtosis measures to test whether market-model residuals for industry portfolios are multivariate normal. He finds strong evidence against this hypothesis. In contrast, Affleck-Graves and McDonald (1989) look at the properties of the marginal distribution of the time series of individual residuals. For example, with respect to market-model residuals of size portfolios over 5-year monthly sample periods, they find that in the prewar period about half of the residuals are significantly different from the normal distribution at the 5% level. In the postwar period, only 15%–20% are significant. These results are difficult to interpret, especially given the well-documented high correlation across portfolios. Below, we incorporate the correlation across the portfolio returns in applying the tests of Sections II and III to the question of whether the market-model residuals are MVN distributed.

Consider the disturbance term from the market-model equation for excess returns on the 10 size decile portfolios (denote R_{it}):

$$\epsilon_{it} = R_{it} - \alpha_i - \beta_i R_{mt}, \quad i = 1, \ldots, 10,$$

where

$$E[\epsilon_{it}] = 0,$$

$$E[\epsilon_{it} R_{mt}] = 0,$$

and

$$R_{mt} = \text{the excess market return.}$$

The econometrician's goal is to test whether the ϵ_{it}, $i = 1, \ldots, 10$, are MVN distributed. These ϵ_{it}'s, however, are unobservable. Nevertheless, it is possible to test the multivariate normal hypothesis by testing whether the residuals from the market-model regression of R_{mt} on the R_{it}'s conform to an MVN distribution.[6] Using the GMM test procedure in Section III, this test is correctly specified and is equivalent asymptotically to testing the disturbance terms directly. Moreover, under the null hypothesis that the disturbance terms are MVN and under some weak additional assumptions, it is possible to show that the distributional results given in Sections II and III are exactly the same for the residuals (where $\hat{\epsilon}_{it}$ is substituted for R_{it}).[7] Estimation of α_i and β_i, therefore, poses no real efficiency problem asymptotically in the GMM framework.

With respect to the multivariate normality of the market model residuals on the 10 size portfolios, we calculate the skewness, kurtosis, and cross-moments test statistics of Sections II and III. These tests are performed over 5-year monthly periods from 1926 to 1990 and over the overall period. For each measure, we report the number of individual rejections and the Wald test statistic with corresponding p-value. These results are provided in table 3 (i.e., Sec. II tests) and in table 4 (i.e., Sec. III tests).

6. Specifically, the market model residuals equal

$$\hat{\epsilon}_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt}, \quad i = 1, \ldots, 10; \quad t = 1, \ldots, T,$$

where

$$\hat{\alpha}_i = \overline{R}_{it} - \hat{\beta}_i \overline{R}_{mt},$$

$$\hat{\beta}_i = \frac{\displaystyle\sum_{t=1}^{T} (R_{it} - \overline{R}_{it})(R_{mt} - \overline{R}_{mt})}{\displaystyle\sum_{t=1}^{T} (R_{mt} - \overline{R}_{mt})^2},$$

$$\overline{R}_{it} = \frac{1}{T} \sum_{t=1}^{T} R_{it},$$

and

$$\overline{R}_{mt} = \frac{1}{T} \sum_{t=1}^{T} R_{mt}.$$

7. The condition sufficient for asymptotic equivalence is $E[R_m \epsilon_{it}^r \epsilon_{jt}^q] = 0$ for all i and j, and for any r and q dictated via the particular moment conditions being estimated. For example, in the example using kurtosis in Sec. IIA, we must assume that $E[R_{mt} \epsilon_{it}^3] = 0$ for all i. A sufficient condition for this being true is that the disturbance terms are independent of the market return, a common assumption in the literature. Of interest to our earlier analysis, this condition is true if the asset returns and the market return are multivariate normal.

TABLE 3   Multivariate Normal Test: Market-Model Residuals

Period	Rejections (Max = 10)	Wald Test for Skewness	Table p-Value	Monte Carlo p-Value	Rejections (Max = 10)	Wald Test for Kurtosis	Table p-Value	Monte Carlo p-Value
1926–30	2	39.3960	.0000	.0006	3	56.9320	.0000	.0024
1931–35	8	125.9810	.0000	.0000	8	788.6689	.0000	.0000
1936–40	3	96.7740	.0000	.0000	3	447.9109	.0000	.0000
1941–45	4	87.1440	.0000	.0000	8	240.7430	.0000	.0000
1946–50	3	15.7260	.1077	.0852	0	9.1680	.5162	.2626
1951–55	3	25.9710	.0038	.0084	1	11.0580	.3530	.1804
1956–60	6	49.2820	.0000	.0000	6	114.0180	.0000	.0000
1961–65	0	9.6290	.4736	.3932	3	21.9500	.0154	.0428
1966–70	0	11.2000	.3422	.2716	0	9.0710	.5253	.2660
1971–75	9	57.3950	.0000	.0000	8	81.0550	.0000	.0006
1976–80	3	53.0850	.0000	.0000	8	54.8470	.0000	.0024
1981–85	4	30.3940	.0007	.0030	3	32.8140	.0003	.0154
1986–90	0	10.8720	.3676	.2948	0	6.0880	.8078	.4962

NOTE.—This table tests for multivariate normality of the market-model residuals from 10 size portfolios during the period 1926–90. The sample is broken up into nonoverlapping subperiods of 5 years each. Column 1 contains the period over which the multivariate test is performed. Column 2 contains the number of rejections (max = 10) based on the univariate skewness test. Column 3 contains the multivariate Wald test for skewness. Columns 4 and 5 contain the table and Monte Carlo p-values of this statistic. Column 6 contains the number of rejections (max = 10) based on the univariate kurtosis test. Column 7 contains the multivariate Wald test for kurtosis. Columns 8 and 9 contain the table and Monte Carlo p-values of this statistic. Monte Carlo p-values are based on a simulation where test statistics are calculated using 60 observations from a multivariate normal distribution with variance-covariance matrix equal to the empirical variance-covariance matrix of the market-model residuals from 1986–90. Five thousand repetitions of the simulation are made.

TABLE 4   Multivariate Normal Test: Market-Model Residuals

Period	Rejections (Max = 9)	Wald Test for Coskewness	Table p-Value	Monte Carlo p-Value	Rejections (Max = 9)	Wald Test for Cokurtosis	Table p-Value	Monte Carlo p-Value
1926–30	3	31.5720	.0002	.0008	3	31.2380	.0003	.0160
1931–35	8	36.4850	.0000	.0004	8	307.2649	.0000	.0000
1936–40	8	170.3430	.0000	.0000	7	725.6660	.0000	.0000
1941–45	9	85.8070	.0000	.0000	9	421.5071	.0000	.0000
1946–50	8	11.9120	.2183	.1792	0	4.0610	.9073	.7042
1951–55	8	26.0290	.0020	.0044	1	14.6970	.0096	.0832
1956–60	9	74.6160	.0000	.0000	9	240.2690	.0000	.0000
1961–65	0	8.3200	.5022	.4108	1	13.1120	.1576	.1070
1966–70	0	5.9040	.7495	.6632	0	4.9880	.8354	.5612
1971–75	9	39.6560	.0000	.0004	9	75.8460	.0000	.0006
1976–80	7	42.0170	.0000	.0004	9	51.4750	.0000	.0036
1981–85	4	18.6760	.0281	.0316	1	37.4420	.0000	.0096
1986–90	0	14.0120	.1219	.1026	0	3.0070	.9640	.8618

NOTE.—This table tests for multivariate normality of the market-model residuals from 10 size portfolios during the period 1926–90. The sample is broken up into nonoverlapping subperiods of 5 years each. Column 1 contains the period over which the multivariate test is performed. Column 2 contains the number of rejections (max = 9) based on the univariate coskewness test. Column 3 contains the multivariate Wald test for coskewness. Columns 4 and 5 contain the table and Monte Carlo p-values of this statistic. Column 6 contains the number of rejections (max = 9) based on the univariate cokurtosis test. Column 7 contains the multivariate Wald test for cokurtosis. Columns 8 and 9 contain the table and Monte Carlo p-values of this statistic. Monte Carlo p-values are based on a simulation where test statistics are calculated using 60 observations from a multivariate normal distribution with variance-covariance matrix equal to the empirical variance-covariance matrix of the market-model residuals from 1986–90. Five thousand repetitions of the simulation are made. The market-model residuals from size portfolio 1 are used as the reference asset for both the coskewness and cokurtosis test.

The multivariate tests provide very similar results. Nine of the 13 subperiods show a strong rejection of multivariate normality for all four tests. In the prewar period (1926–40), all of the three subperiods reject multivariate normality. In the postwar period (1946–90), multivariate normality is rejected for five of the nine subperiods.

The results highlight some of the pitfalls of reliance on univariate statistics when these statistics are highly related. For example, for the 1926–30 subperiod only two of the 10 skewness measures, three of the 10 kurtosis measures, three of the nine cross-skewness measures, and three of the nine cross-kurtosis measures reject normality. All four of the multivariate tests provide strong evidence against normality. Similarly, for the 1981–85 subperiod only one of the nine cross-kurtosis measures indicates rejection of normality, yet the multivariate test indicates a strong rejection of multivariate normality. In contrast, for the 1946–50 subperiod eight of the nine cross-skewness tests reject normality, yet the multivariate tests do not reject normality.

V.   Extensions

A.   Relaxation of i.i.d. Assumption

The multivariate test procedure in Section III assumed that $\{R_t\}_{t=1}^T$ are drawn from an i.i.d. multivariate distribution. In terms of the distributional results, however, the test procedure requires only that the R_t be stationary and ergodic and that the moment restrictions have a finite variance-covariance matrix. When the i.i.d. assumption is relaxed, for the analysis to make sense, the form of serial dependence (e.g., serial correlation, conditional heteroscedasticity, etc.) must be internally consistent with the multivariate normal null hypothesis. For example, suppose stock returns are serially correlated following some autoregressive integrated moving average (ARIMA) model. Then if the innovations across returns each period are i.i.d. multivariate normal, the returns will also be multivariate normally distributed. The test statistics, however, will be misspecified because the estimator for S_0 will no longer be consistent. The econometrician is then confronted with several issues. First, under the more general framework, is the methodology described in Section III still valid? Second, if the methodology is valid, is it necessary that we know what precise process the R_t follow? Finally, how should the tests be performed?

Without loss of generality, suppose that R_{it} follows an autoregressive (AR) process of order 1. Further, assume that the econometrician is interested in estimating and testing properties of the variance $\sigma_{R_i}^2$. First, note that the best estimator (in terms of minimizing the asymptotic variance) for $\sigma_{R_i}^2$ is the ML estimator, $(\sigma^{MLE})^2 = (\sigma_{\epsilon}^{MLE})^2/[1 -$

$(\gamma^{\text{MLE}})^2]$, where σ_ϵ^2 is the variance of the market-model innovation term and γ is the AR parameter. It is possible to show that this estimator is, for all relevant asymptotic comparisons, equal to the sample variance, $\hat{\sigma}_{R_i}^2$. Thus, as in Section III, the optimal GMM estimators are still the sample means, variances, and correlations. Intuitively, since one cannot do better asymptotically than MLE, the GMM estimation will again always pick these sample moments in estimation regardless of higher-order moment restrictions or of any serial correlation in the returns.

However, even though the GMM estimates are the same under serial dependence, this is not true of the estimators' asymptotic variance-covariance matrix. For example, in the AR(1) case, it is possible to show that the variance of $\hat{\sigma}_{R_i}^2$ is $[2\sigma_{R_i}^4(1 + \gamma^2)]/(1 - \gamma^2)$. Clearly, the S_0 estimator for the i.i.d. case (i.e., $2\hat{\sigma}_{R_i}^4$) is then not consistent under this more general AR formulation. Clearly, if the econometrician knows the order of the AR process, then the estimation can be performed directly. For example, in the AR(1) example, the asymptotic variance of $\hat{\sigma}_{R_i}^2$ can be estimated consistently via $[2\hat{\sigma}_{R_i}^4(1 + \hat{\gamma}^2)]/(1 - \hat{\gamma}^2)$, where $\hat{\gamma}$ is the first-order autocorrelation of R_i.

Knowledge of the precise order of the process is usually not known a priori. However, several procedures for estimating the variance-covariance matrix in the presence of unknown serial dependence have been developed. For example, Hansen (1982) shows how to adjust S_0 to reflect general dependence:

$$S_0 = \sum_{l=-\infty}^{l=+\infty} E[h(R_t)h(R_{t-l})'].$$

One such estimator for S_0 is the sample moment estimate of S_0, truncated at some "reasonable" value for l (see Hansen and Singleton [1982] for a discussion). Unfortunately, this estimator is not assured of being positive definite. Nevertheless, there are similar autocorrelation consistent estimators that ensure positive definiteness (see, e.g., Newey and West 1987; and Andrews 1991).

In summary, the GMM procedure in the presence of serial dependence involves three steps:

First, apply the GMM methodology to get the optimal estimates. In the case of the MVN distribution, GMM produces the sample moments, $\hat{\mu}_R$, $\hat{\sigma}_R$, and $\hat{\rho}_{ij}$, for all $i \neq j$.

Second, estimate S_0 using an autocorrelation consistent estimator. Denote this estimator $\hat{S}_T^{ac}$.

Third, test the MVN distribution by calculating the statistic

$$J_T \equiv Tg_T(\hat{\theta})'S_T^{ac-1}g_T(\hat{\theta}),$$

and then evaluating J_T at the appropriate level of significance.

B. The Optimal Test

Small sample considerations aside, the GMM test procedure outlined above generates an infinite number of test statistics, all with asymptotic χ^2 distributions. In practice, this class of statistics is limited by the number of assets and by a finite number of restrictions because we have only a limited number of time-series observations at our disposal. Nevertheless, the class of statistics is significantly larger than the skewness/kurtosis statistics studied in Sections II and III. Given that the econometrician has numerous moment conditions available to him when testing the MVN distribution's joint restrictions, a natural question is, Which moment conditions should he choose in estimation?

Sometimes, of course, the choice of restrictions comes quite naturally; for example, the econometrician may wish to focus on particular moments such as skewness or kurtosis. As mentioned above, there are a number of economic reasons why the finance literature has focused on these moments. Similarly, the theory may lead to particular moment conditions; an illustration of this is the Kraus and Litzenberger (1976) three-moment capital asset pricing model (CAPM), in which skewness plays a substantive role. These criteria, as with any choice of test statistics, are somewhat subjective. In terms of a more objective criterion, which moment restrictions should be chosen?

1. *The approximate-slope procedure.*   Suppose we fix the number of restrictions being tested at $R - M = Q^*$. Irrespective of whether the test statistic takes the form of equation (1) or the $J_T(\hat{\theta})$ statistic in Section III, the asymptotic distribution of the statistic is $\chi^2_{Q^*}$. From an asymptotic point of view, there is no difference between choosing a particular set of Q^* overidentifying moment restrictions over an alternative set of Q^* restrictions under the null hypothesis that asset returns are MVN distributed.[8] There is a difference asymptotically, however, between these statistics under specified alternative multivariate distributions, that is, with respect to their relative asymptotic power. Using Bahadur's (1960) concept of approximate slope, Geweke (1981) develops a procedure for comparing the asymptotic power of test statistics by comparing their approximate slopes.[9]

8. There may be a difference in small samples, however. To the extent that the goal of asymptotic theory is to approximate the small sample distribution, one can argue we should choose the moment restrictions which best fit the $\chi^2_{Q^*}$ distribution. Although there is little theory suggesting which moments to choose on this basis, it is well known that higher-order moments provide the poorest approximations in small samples. One would suspect, therefore, that the best approximation occurs with lower-order moments and cross moments. This issue deserves considerable more study but is beyond the scope of this article. (See Serfling [1980] for a discussion of biases in sample moments and Mardia [1980] for convergence properties of some moments in the case of normality.)

9. The approximate slope is a measure of the rate at which the null hypothesis becomes more incredible as the sample size increases. Specifically, for a given alternative and fixed power, $-2 \ln(\alpha)/T$ converges almost surely to the approximate slope of the test, where α is the marginal significance level of the test.

Of special interest to this article, Geweke (1981) proves two important results. First, if the test statistic has an asymptotic χ^2 distribution, then the approximate slope equals the probability limit (plim) of the statistic deflated by sample size. Second, for a fixed number of restrictions Q^*, the ratio of approximate slopes between two test statistics will equal the inverse ratio of the minimum number of observations needed to achieve a given power (i.e., as we let the size of the nonrejection region get arbitrarily large). For example, a statistic with one-half the approximate slope of another statistic will need roughly twice as many observations to reject the MVN distribution.

Therefore, under a given alternative multivariate distribution and under a fixed number of restrictions Q^*, one objective criterion for choice of test restrictions is to pick the moment conditions that maximize the approximate slope of the test. This is an especially appealing method because the result will often not depend on nuisance parameters. That is, similar to the asymptotic null distribution being derived for arbitrary μ and σ, the approximate-slope results will also be independent of the values of μ and σ. This will not necessarily be true of power calculations based on Monte Carlo simulations. The drawback of the approximate-slope procedure relative to Monte Carlo simulation, however, is that it is valid only asymptotically.

Nevertheless, to see this procedure in practice, consider the $J_T^i(\hat{\theta})$ statistic, where i represents just one set of particular moment restrictions. In terms of $J_T^i(\hat{\theta})$, choose the restrictions i that maximize

$$\text{plim}[g_T^i(\hat{\theta})' S_0^{i-1} g_T^i(\hat{\theta})]$$

under a given alternative. At first glance, this task may seem somewhat daunting. But, in fact, these probability limits are fairly straightforward to calculate. This is because $\text{plim}[g_T(\hat{\theta})]$ and S_0 are simply moments of the distribution, which (if they exist) can be calculated directly under the alternative distribution. These calculations are made even easier when we realize that under the MVN null hypothesis our estimate of S_0 is a known function of only the means, variances, and correlations between the assets. As long as we impose the null hypothesis in estimation, all that we need to calculate, therefore, is plim $[g_T(\hat{\theta})]$ under the alternative distribution.

2. *Example.*   As an illustration of the approximate-slope procedure, consider the "kurtosis-based" statistics of Sections II*B* and III*A*. Since these statistics have asymptotic χ^2 distributions, their approximate slopes under given alternatives can be compared directly. One popular alternative distributional model to multivariate normality is the multivariate Student t (see Blattberg and Gonedes [1974] for an example of empirical work and Ingersoll [1987] for some theoretical justification of this distribution). Under the multivariate t assumption,

the plim of K_i and K_{ij} are readily calculated (see Johnson and Kotz [1970] and Zellner [1971] for distributional properties of the multivariate t).

For simplicity, consider testing just one restriction and choosing between the kurtosis and cross-kurtosis tests, that is, $K_i = 0$ versus $K_{ij} = 0$. Using the asymptotic distributional results in Sections II and III above, the approximate slopes of K_i (denote c_i) and K_{ij} (denote c_{ij}) are given by

$$c_i = \frac{3}{2(\nu - 4)^2}$$

and

$$c_{ij} = \left(1 + \frac{3\rho_{ij}^4}{\rho_{ij}^4 + 4\rho_{ij}^2 + 1}\right)\frac{1}{(\nu - 4)^2},$$

where ν = degrees of freedom parameter for multivariate t. Note that the approximate slopes depend only on the degrees of freedom parameter ν and (in the case of cross kurtosis) the true correlation between asset returns. The approximate slope, and thus the asymptotic power, of the tests decreases as ν increases. This is expected as Student t with high ν is approximately normal. The ratio of the approximate slopes, c_i/c_{ij}, takes on an especially interesting form. It is independent of ν (and, therefore, holds for all multivariate t alternatives) and depends only on the true correlation between asset returns. In particular, c_i/c_{ij} ranges from 1 to 1.5 as a decreasing function of the correlation between the assets, ρ_{ij}. As an illustration, suppose the correlation between the two asset returns is 80%. In this case, c_i is 31.96% greater than c_{ij}. Therefore, if we decide to use the cross-kurtosis test instead of the more standard kurtosis measure, we will need almost one-third more observations to achieve the same power in testing multivariate normality against any multivariate Student t alternative.

These results do not hold generally for the multiple-restriction case. The ratio of the approximate slope of the Wald test for the kurtosis measures over the approximate slope of the Wald test for the cross-kurtosis measures can be either greater than or less than one. This ratio depends on the correlation matrix across asset returns. As an example, consider fixing the number of restrictions at nine (with size decile 1 as the benchmark asset) and comparing the kurtosis Wald statistic and cross-kurtosis Wald statistics of Sections II and IIIA, respectively. In order to compare their approximate slopes, it is necessary to specify the complete correlation matrix of the asset returns. Suppose the true correlation structure is equal to the sample cross correlations of the market-model residuals for the 1986–90 subperiod. The resulting ratio of approximate slopes, c_i/c_{ij}, is 1.266. Evidently,

in this particular case, the Wald statistic for the kurtosis measures provides greater asymptotic power.

C. On Imposing the Null Distribution in Estimation

So far, we have imposed the null distribution when calculating the asymptotic variance-covariance matrix of the estimators. In particular, our method calculates the variance-covariance matrix analytically under the null hypothesis that $R_t \overset{d}{\sim} F(R, \theta)$. An estimate of $S_0(\theta)$ can then be provided through $S_0(\bar{\theta})$, where $\bar{\theta}$ is a consistent estimate of θ (e.g., one possible choice being the sample means, variance, and cross correlations). An alternative estimation strategy involves not imposing the null distribution and calculating the variance-covariance matrix using sample estimates. Consider the estimator, $S_T(\bar{\theta})$, where $\bar{\theta}$ is a consistent estimate of θ and S_T is the sample estimate of the variance-covariance matrix of the moment conditions. It is possible to show that $S_T(\bar{\theta})$ is consistent and has asymptotically equivalent properties to $S_0(\bar{\theta})$ under the null hypothesis.

What are possible reasons for choosing one estimator over another? In terms of the size of the GMM test, the analytical estimator ($S_0(\bar{\theta})$) requires estimation of only the first and second moments of the distribution. This has two benefits in small samples. First, there are well-known problems with estimating higher-order moments in small samples—biases and slow convergence are prevalent (see n. 8 above). In contrast to $S_0(\bar{\theta})$, the sample estimator ($S_T(\bar{\theta})$) requires estimating twice the highest-order moment restriction. For example, a kurtosis-based test requires estimation of the eighth moment in addition to lower-order moments. This points to the second benefit of using $S_0(\bar{\theta})$ over $S_T(\bar{\theta})$. The estimator $S_T(\bar{\theta})$ requires estimation of many more moments. For example, with the moment conditions given in equation (5), $S_0(\bar{\theta})$ estimates five moments, while $S_T(\bar{\theta})$ needs to estimate $[N(N-1)]/2 = 91$ moments. The sampling errors in small samples associated with repeated use of the data (in this case, multiple-moment estimation) is therefore much worse for the estimator $S_T(\bar{\theta})$.

In terms of the power of the test, there is no reason a priori to choose one estimator over the other. Note that $S_0(\bar{\theta})$'s estimate is the same irrespective of whether the data come from the null or alternative. In contrast, the estimator $S_T(\bar{\theta})$ picks up information contained in the alternative distribution. Therefore, under some alternative distribution, these different estimators will provide quite different weights on the moment restrictions not set equal to zero. The ultimate choice then depends on the class of alternatives the econometrician deems reasonable.

One way to address this issue formally is via the approximate slope procedure discussed in Section V*B* above. Fixing the alternative, the approximate slope procedure can then be used to choose ex ante

the appropriate estimate for S_0 in terms of power. In particular, under the alternative distribution and a fixed number of restrictions, $\text{plim}[g_T(\hat{\theta})'S_0^{-1}g_T(\hat{\theta})]$ can be compared to $\text{plim}[g_T(\hat{\theta})'S_T^{-1}g_T(\hat{\theta})]$. If, for example, the ratio of approximate slopes $\text{plim}[g_T(\hat{\theta})'S_0^{-1}g_T(\hat{\theta})]/$ $\text{plim}[g_T(\hat{\theta})'S_T^{-1}g_T(\hat{\theta})]$ is two-to-one, then the sample estimation procedure (using $S_T(\bar{\theta})$) has less asymptotic power, requiring about twice as many observations to be equivalent.

One final comment pertains to the estimation procedure in general. Suppose we wish to construct confidence intervals around the statistics that are valid under many distributional alternatives. As an example, consider the following sample moments for univariate kurtosis restrictions for R_i and R_j:

$$
g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}
\begin{vmatrix}
R_{it} - \mu_i \\
R_{jt} - \mu_j \\
(R_{it} - \mu_i)^2 - \sigma_i^2 \\
(R_{jt} - \mu_j)^2 - \sigma_j^2 \\
(R_{it} - \mu_i)^4 - \sigma_i^4(3 + K_i^*) \\
(R_{jt} - \mu_j)^4 - \sigma_j^4(3 + K_j^*)
\end{vmatrix}. \tag{6}
$$

Construction of "distribution-free" confidence intervals around $\hat{K}_i^*$ and $\hat{K}_j^*$ is straightforward. The steps can be described as follows. First, the econometrician calculates both the sample derivative matrix, $D_T = 1/T \sum_{t=1}^{T} \partial h(R_t, \hat{\theta})/\partial\theta$, and the sample variance-covariance matrix, $S_T = 1/T \sum_{t=1}^{T} h(R_t, \hat{\theta})h(R_t, \hat{\theta})'$. Using these estimates, the asymptotic variance of the estimators can be consistently estimated by $[D_T'S_T^{-1}D_T]^{-1}$. The standard errors around $\hat{K}_i^*$ and $\hat{K}_j^*$, therefore, do not depend explicitly on an imposed null distribution. Tests for an MVN distribution (i.e., $K_i^* = K_j^* = 0$), tests for a Poisson distribution (i.e., $K_i^* = 1/\mu_i$ and $K_j^* = 1/\mu_j$), and so forth, can be readily performed.

The obvious benefit behind this type of generalization is that the estimators and corresponding confidence intervals are robust to many distributions (i.e., to ones with well-defined moments). The drawback is that the properties of these statistics in small samples may be suspect for the reasons described above.[10]

---

10. A similar kind of analysis can be performed for overidentifying restrictions, although some null (no matter how weak) needs to be imposed via the moment restrictions. For example, one might wish to test whether the data come from a class of multivariate symmetric distributions in which skewness is zero. The procedure here is to first estimate consistent estimates, $\hat{\theta}$, of the parameters using some weighting matrix (e.g., the identity matrix I):

$$
\min_{\theta} g_T(\theta)'Ig_T(\theta).
$$

VI. Conclusion

It is difficult to interpret individual test statistics for univariate normality in stock returns across assets. By explicitly taking into account the contemporaneous correlation between asset returns, it is possible to jointly test the hypothesis that stock returns are normally distributed. These tests restrict themselves, however, to investigations of the marginal distributions of the assets—departures from multivariate normality may be more prevalent in the joint distribution of the assets. This article proposes a class of tests that exploits information contained in both the marginal and joint moments of asset returns. Of statistical interest, this class of test statistics is easy to calculate with well-known asymptotic distributions. As a technical by-product, we discuss a procedure for evaluating the most powerful statistic within this class.

In applying these tests to stock returns and market-model residuals, we find highly significant evidence of nonnormality in both the marginal and joint distributions of these variables. At least empirically, therefore, the multivariate normal assumption cannot be justified. With respect to alternative multivariate distributions (e.g., such as the multivariate *t*), the techniques introduced here can be used to test the appropriateness of these alternative distributions. To this extent, this article should have applications elsewhere in the literature.

Appendix

Dow Jones Firms (January 1951–June 1968)

Allied Corporation	(ALD)
Alcoa	(AA)
American Tobacco	(AMB)
AT & T	(T)
Anaconda Co.	(A)
Bethlehem Steel Corporation	(BS)
Standard Oil (Calif.)	(CHV)
Chrysler Corp	(C)
Du Pont E. I. De Nemours & Company	(DD)
Eastman Kodak Company	(EK)

In the second step, the econometrician calculates an optimal weighting matrix using these consistent estimates, denoted $S_T(\hat{\theta})$. The final step then has the econometrician calculate the optimal GMM estimates, $\hat{\theta}$, from

$$\min_{\theta} g_T(\theta)' S_T(\bar{\theta})^{-1} g_T(\theta),$$

and the corresponding overidentifying restrictions test statistic using these optimal estimates:

$$T g_T(\hat{\theta})' S_T(\bar{\theta})^{-1} g_T(\hat{\theta}).$$

Swift & Co.   (ESM)
Standard Oil (N.J.)   (XON)
General Electric Company   (GE)
General Foods Corporation   (GF)
General Motors Corporation   (GM)
Goodyear Tire & Rubber Company   (GT)
International Nickel   (N)
International Paper   (IP)
Johns Manville Corporation   (JM)
International Harvester   (NAV)
Owens-Illinois, Inc.   (OI)
American Can   (PA)
Procter & Gamble Company   (PG)
Sears Roebuck & Company   (S)
Texaco Inc.   (TX)
United States Steel Corporation   (X)
Union Carbide Corporation   (UK)
United Aircraft   (UTX)
Westinghouse Electric Corporation   (WX)
F. W. Woolworth Company   (Z)

References

Affleck-Graves, J., and McDonald, B. 1989. Nonnormalities and tests of asset pricing theories. *Journal of Finance* 44:889–908.

Andrews, Donald. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59:817–58.

Bahadur, R. R. 1960. Stochastic comparison of tests. *Annals of Mathematical Statistics* 31:276–95.

Blattberg, R., and Gonedes, N. 1974. A comparison of the stable and student distributions as stochastic models for stock prices. *Journal of Business* 47:244–80.

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

Bookstaber, R., and McDonald, J. B. 1987. A general distribution for describing security price returns. *Journal of Business* 60:401–24.

Breeden, D. 1986. Consumption, production, inflation and interest rates: A synthesis. *Journal of Financial Economics* 16:3–40.

Clark, Peter K. 1973. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41:135–56.

Cox, D., and Small, N. J. 1978. Testing multivariate normality. *Biometrika* 65:263–72.

Fama, Eugene. 1965. The behavior of stock market prices. *Journal of Business* 38:34–105.

Fama, E. 1976. *Foundations of Finance*. New York: Basic.

Geweke, J. 1981. The approximate slope of econometric tests. *Econometrica* 49, no. 6 (November): 1427–42.

Gibbons, M.; Ross, S.; and Shanken, J. 1989. A test of the efficiency of a given portfolio. *Econometrica* 57:1121–52.

Hansen, L. 1982. Large sample properties of generalized method of moment estimators. *Econometrica* 50, no. 4 (July): 1029–54.

Hansen, L., and Singleton, K. 1982. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50 (September): 1269–86.

Harris, Lawrence E. 1986. Cross-security tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis* 21:39–46.

Ingersoll, J. 1987. *Theory of Financial Decision Making*. Totowa, N.J.: Rowman & Littlefield.

Johnson, N., and Kotz, S. 1970. *Distributions in Statistics: Continuous Multivariate Distributions*. New York: Wiley.

Kraus, A., and Litzenberger, R. 1976. Skewness preference and the valuation of risky assets. *Journal of Finance* 31:1085–1100.

Lo, Andrew W., and MacKinlay, A. Craig. 1988. Stock prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies* 1:41–66.

MacKinlay, C. 1985. An analysis of multivariate financial tests. Ph.D. dissertation, University of Chicago.

MacKinlay, C. 1987. On multivariate tests of the CAPM. *Journal of Financial Economics* 18:341–71.

MacKinlay, C., and Richardson, M. 1991. Using generalized method of moments to test mean-variance efficiency. *Journal of Finance* 46:511–27.

Mardia, K. 1970. Measures of multivariate skewness and kurtosis with applications. *Biometrika* 57:519–30.

Mardia, K. 1980. Tests of univariate and multivariate normality. In *Handbook of Statistics*. Vol. 1, *Analysis of Variance*. Amsterdam: North Holland.

Newey, W., and West, K. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (May): 703–8.

Officer, R. 1971. A time series examination of the market factor of the New York Stock Exchange. Ph.D. dissertation, University of Chicago.

Schwert, G. William. 1989. Why does stock market volatility change over time? *Journal of Finance* 44:1115–54.

Serfling, R. 1980. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.

Tauchen, George E., and Pitts, Mark. 1983. The price variability-volume relationship on speculative markets. *Econometrica* 51:485–505.

Zellner, A. 1971. *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley.

Zhou, Ghoufu. 1991. Asset pricing tests under alternative distributions. Working paper, Washington University at St. Louis.