Chi-Square Tests For Multivariate Normality With Application To Common Stock Prices

by

David S. Moore and John B. Stubblebine

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series #78-17

July 1978

Research supported in part by the Air Force Office of Scientific Research under Grant AFOSR 77-3291.

Chi-Square Tests For Multivariate Normality With Application To Common Stock Prices

David S. Moore and John B. Stubblebine*

The theory of chi-square tests with data-dependent cells is applied to provide tests of fit to the family of p-variate normal distributions. The cells are bounded by hyperellipses $(x-\overline{X})'S^{-1}(x-\overline{X}) = c_i$ centered at the sample mean $\overline{X}$ and having shape determined by the sample covariance matrix S. The Pearson statistic with these cells is affine-invariant, has a null distribution not depending on the true mean and covariance, and has asymptotic critical points between those of $\chi^2(M-1)$ and $\chi^2(M-2)$ when M cells are employed. The test is insensitive to lack of symmetry, but peakedness, broad shoulders and heavy tails are easily discerned in the cell counts. Multivariate normality of logarithms of relative prices of common stocks, a common assumption in financial markets theory, is studied using the statistic described here and a large data base.

KEY WORDS: Chi-square tests, Goodness of fit, Multivariate normal distribution, Distribution of stock prices.

1. INTRODUCTION

Despite the frequency with which multivariate normality is assumed in multivariate statistical models, the literature on tests of fit for the multivariate normal family is much smaller than that concerned with univariate normality. Andrews, Gnanadesikan and Warner (1973) provide a summary of proposed methods and also offer comments on the process of assessing multivariate normality. They note that because of the variety of possible departures from joint normality it is desirable to have methods sensitive to different types of departure rather than seeking a single best method. Aside from application of univariate techniques to marginal data, the techniques discussed by these authors are categorized as (1) transformation-related methods; (2) tests based on distributional densities; (3) tests based on unidimensional views of multivariate data. This paper studies a chi-square test for multivariate normality, using data-dependent cells bounded by hyperellipses which are surfaces of constant probability density function for the normal distribution with parameters estimated from the data. That is, the cell boundaries are concentric hyperellipses $(x-\overline{X})'S^{-1}(x-\overline{X}) = c_i$ centered at the sample mean $\overline{X}$ and with shape determined by the sample covariance matrix S. The figure illustrates these cells in the bivariate case.

This test is a member of the second class of techniques in the categorization of Andrews, et al. It has several advantages. First, it is easy to choose cells having prespecified estimated cell probabilities, and relatively easy to evaluate the test statistic. Second, the large sample theory of the test is nearly standard, and allows use of chi-square critical points to assess the significance of the statistic. Third, the nature of departures from normality is indicated by the observed cell frequencies. This kind of data analysis is particularly simple when the cells are equiprobable. Such common departures from normality as peakedness, broad shoulders, and heavy tails are directly apparent in the cell counts. The first and third classes of techniques discussed by Andrews, et.al. rarely give so clear an indication of the type of departure present. On the other hand, the cell counts and tests based on them are not sensitive to lack of symmetry. Moreover, though assessment of the power of tests of fit is even more ambiguous here than in the univariate case, tests of chi-square type are rarely characterized by outstanding overall sensitivity. The proposed test must compete for use, usually as one of several techniques applied in a data-analytic setting, on the basis of the ease of interpretation provided by the cell counts and the simple large-sample distribution theory.

Section 2 discusses the necessary background in the theory of general chi-square statistics, then describes the proposed test and its large-sample theory. Some lengthy calculations required by this work are deferred to an appendix. This section discusses separately chi-square tests for bivariate circular normality. This case is of special interest because of its prominence in targeting problems and because algebraically simple results can be obtained from our general approach. Section 3 applies the techniques of Section 1 to assess multivariate normality of logarithms of quotients of common stock prices in a large data base. This example is chosen because of its importance in the theory of financial markets, and because previous discussion (mainly univariate) has defended the symmetry of these distributions but hinted at broad shoulders and heavy tails, for which our test is particularly effective. Such important considerations as the choice of cells, checking for lack of symmetry, and assessing the magnitude (as opposed to the significance) of the observed departure from normality are discussed in the context of this application.

In this paper standard notation is used for convergence in law and in probability. The p-variate normal distributions are denoted by $N_p(\mu,\Sigma)$ and the (central) chi-square distributions by $\chi^2(k)$. All vectors are column vectors, with prime denoting transposition. The rank of a matrix A is r(A), and the determinant is $|A|$.

2. TESTING MULTIVARIATE NORMALITY

Independent and identically distributed p-variate random variables $X_1,\ldots,X_n$ are observed. It is desired to test whether the distribution of X_j belongs to the family of p-variate normal distributions $N_p(\mu,\Sigma)$ for some μ and nonsingular Σ. This is a parametric family parametrized by θ consisting of the p means μ_k and the p(p+1)/2 upper triangular entries σ_{jk} of Σ. The density function of X_j under the null hypothesis is

$$f(x|\theta) = (2\pi)^{-p/2}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}. \qquad (2.1)$$

The maximum likelihood estimators (MLE's) of μ and Σ are the vector $\overline{X}$ of sample means and the matrix S of sample covariances (dividing by n rather than n-1, though for large-sample considerations this is irrelevant).

For a given θ (that is, for given μ and Σ) and $0 = c_0 < c_1 < \ldots < c_M = 1$, define cells

$$E_i(\boldsymbol{\theta}) = \{x \text{ in } R^p\colon c_{i-1} \le (x-\mu)'\Sigma^{-1}(x-\mu) < c_i\}. \qquad (2.2)$$

The surfaces bounding these cells are level surfaces of the density function (2.1). Estimating θ by the MLE $\hat{\theta}_n$, we obtain M data-dependent cells $E_{in} = E_i(\hat{\theta}_n)$, $i=1,\ldots,M$. The use of data-dependent cells in chi-square tests was first studied by Roy (1956) and Watson (1957, 1958, 1959). A general large-sample theory for such tests is given by Moore and Spruill (1975). The tests proceed by treating the cells E_{in} as though they were fixed. Thus the "cell probability" for E_{in} under $f(\cdot|\theta)$ is

$$p_{in}(\theta) = \int_{E_{in}} f(x|\theta)dx. \qquad (2.3)$$

If N_{in} denotes the number of $X_1,\ldots,X_n$ falling in E_{in}, and $\hat{p}_{in}=p_{in}(\hat{\theta}_n)$ is the corresponding estimated cell probability, general chi-square statistics are nonnegative definite quadratic forms in the standardized cell frequencies $[N_{in}-n\hat{p}_{in}]/[n\hat{p}_{in}]^{\frac{1}{2}}$. If V_n denotes the M-vector of standardized cell frequencies, the Pearson chi-square statistic is

$$X^2(\hat{\theta}_n) = V_n'V_n = \sum_{i=1}^{M} \frac{[N_{in}-n\hat{p}_{in}]^2}{n\hat{p}_{in}} .$$

This is not the classical Pearson-Fisher statistic, because θ is estimated by the raw-data MLE $\hat{\theta}_n$. It is rather the data-dependent cell version of the statistic investigated by Chernoff and Lehmann (1954). Although $X^2(\hat{\theta}_n)$ is the statistic of greatest interest, other quadratic forms in V_n may be useful. One such is introduced at (2.9) below.

The present choice of cells, in addition to being natural for the multivariate normal family, has some desirable properties. First, (2.2) and (2.3) show that $\hat{p}_{in}$ is the same as the probability that $(Y-\bar{X})'S^{-1}(Y-\bar{X})$ falls between c_{i-1} and c_i, where $\bar{X}$, S are considered fixed and Y has the $N_p(\bar{X},S)$ distribution. This is the probability assigned to the interval $[c_{i-1},c_i)$ by the $\chi^2(p)$ distribution. So the estimated cell probabilities $\hat{p}_{in}$ are fixed once the c_i are chosen. In particular, M cells equiprobable under the estimated parameter value are obtained by choosing c_i as the i/M point of the $\chi^2(p)$ distribution. This choice is employed in Section 3. The elliptical shape of the level surfaces is not unique to the normal family; Press (1972b) points out that all multivariate symmetric stable densities of order 1 have such level surfaces. Just as with univariate chi-square tests, it is the $\hat{p}_{in}$ that primarily reflect the null hypothesis.

The second desirable property of the proposed test statistics is that they are unaffected by affine transformations of the X_j. The tie between the cells and the data implies affine invariance of V_n and hence of $X^2(\hat{\theta}_n)$. When quadratic forms in V_n other than the Pearson sum of squares are employed, affine invariance depends also on the matrix of the quadratic form. All statistics in this paper are affine-invariant. It follows that the null distribution of our statistics is the same for all μ and Σ, a property not shared by fixed cell chi-square statistics.

The large-sample theory of $X^2(\hat{\theta}_n)$ and other chi-square statistics follows from that of V_n. Moore and Spruill (1975) show that V_n has the same asymptotic distribution under $f(\cdot|\theta_0)$ as if the fixed cells $E_i(\theta_0)$ to which the E_{in} converge had been used. To describe this distribution, and an important difference between the present case and the usual situation, additional notation is required.

Let $J(\theta)$ be the Fisher information matrix for the $N_p(\mu,\Sigma)$ family. If $m=p+p(p+1)/2$ is the dimension of θ, $J(\theta)$ is the $m \times m$ matrix

$$J(\theta) = \left[\begin{array}{c|c} \Sigma^{-1} & 0 \\ \hline 0 & Q^{-1} \end{array}\right] \tag{2.4}$$

where Q is the $p(p+1)/2 \times p(p+1)/2$ covariance matrix of the entries of $\sqrt{n}\, S$. The entries of nQ in terms of the σ_{jk} are given on p. 107 of Press (1972a). Whenever Σ is nonsingular, so also is $J(\theta)$ for the corresponding θ.

For specified θ_0, define

$$p_i(\theta,\theta_0) = \int_{E_i(\theta_0)} f(x|\theta)dx, \qquad (2.5)$$

a notation which makes explicit the dependence of the cell probabilities on both the cells and the parameter value. Define $B(\theta,\theta_0)$ as the $M \times m$ matrix with (i,j)th entry

$$p_i(\theta,\theta_0)^{-\frac{1}{2}} \frac{\partial p_i(\theta,\theta_0)}{\partial \theta_j} .$$

Under regularity conditions which hold in the present case, it follows (Moore and Spruill 1975, Theorem 4.2) that under $f(\cdot|\theta_0)$

$$\mathcal{L}\{V_n\} \to N_M(0, I-qq'-BJ^{-1}B') \qquad (2.6)$$

where $J=J(\theta_0)$, $B=B(\theta_0,\theta_0)$ and q is the M-vector with entries $p_i(\theta_0,\theta_0)^{\frac{1}{2}}$. If c_i is the i/M point of the $\chi^2(p)$ distribution, then $p_i(\theta_0,\theta_0)\equiv 1/M$, thus simplifying both q and B. Because of (2.6), the asymptotic null distribution of $\chi^2(\hat{\theta}_n)$ is determined by the characteristic roots of $I-qq'-BJ^{-1}B'$. The entries of B for any θ_0 (corresponding to μ_0 and Σ_0) are determined from the following lemma. The proof appears in the Appendix.

<u>Lemma 1</u>: When $p_i(\theta,\theta_0)$ is given by (2.1), (2.2), and (2.5), then for any choice of the c_i and any θ_0,

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial \mu_j}\right|_{\theta=\theta_0} = 0 \qquad 1 \le i \le M,\ 1 \le j \le p$$

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial \sigma_{jk}}\right|_{\theta=\theta_0} = d_i\sigma^{jk} \qquad 1 \le i \le M,\ 1 \le j \le k \le p$$

where σ^{jk} are the entries of Σ_0^{-1} and

$$d_i = (c_{i-1}^{p/2}\, e^{-c_{i-1}/2} - c_i^{p/2}\, e^{-c_i/2})b_p/2$$

$$b_p = [p(p-2)\ldots(4)(2)]^{-1} \qquad p \text{ even}$$

$$= (2/\pi)^{\frac{1}{2}}[p(p-2)\ldots(5)(3)]^{-1} \qquad p \text{ odd.}$$

In the usual situation (Chernoff and Lehmann 1954, Watson 1959, Moore and Spruill 1975) the matrix B, and hence also $BJ^{-1}B'$, has rank m. Lemma 1 shows that in the present case all columns of B are scalar multiples of the vector $(d_1,\ldots,d_M)'$ so that $r(B)=1$. When $r(B)=m$, the characteristic roots of $I-qq'-BJ^{-1}B'$ are M-m-1 1's, one 0 and $1-\delta_i$ where δ_i are the m nonzero roots of $BJ^{-1}B'$ and satisfy $0 \le \delta_i < 1$ provided that J is nonsingular. (See e.g. Watson 1958, pp. 51-54.) Inspection shows that the same arguments apply when $r(B)=k < m$ and k replaces m throughout. That is, the rank of B rather than the number of unknown parameters is the essential interpretation of m in the usual case. In the present case k=1. When equiprobable cells are employed, the nonzero characteristic root of $BJ^{-1}B'$ is shown in the Appendix to be $\delta=2Mp\Sigma_1^M d_i^2$. We have now established the following result.

Theorem 1: Under the null hypothesis of normality the limiting distribution of the Pearson statistic $X^2(\hat{\theta}_n)$ for cells E_{in} with parameters estimated by the MLE's $\overline{X}$ and S is the distribution of

$$\chi^2(M-2) + \lambda\, \chi^2(1) \qquad (2.7)$$

where $\chi^2(M-2)$ and $\chi^2(1)$ are independent chi-square random variables with the indicated degrees of freedom and $0 < \lambda \le 1$. When $\hat{p}_{in} \equiv 1/M$, then

$$\lambda = 1-2Mp\sum_{i=1}^{M} d_i^2. \qquad (2.8)$$

Theorem 1 implies that the asymptotic critical points of $X^2(\hat{\theta}_n)$ fall between those of the $\chi^2(M-2)$ and $\chi^2(M-1)$ distributions. Unless M is very small, these bounds are sufficient for use in practice. This consequence of the symmetry of the cells chosen contrasts markedly with the $r(B)=m$ case. Then, as Chernoff and Lehmann first noted, asymptotic critical points of the Pearson statistic with raw-data MLE's fall between those of $\chi^2(M-m-1)$ and $\chi^2(M-1)$. Since here $m=p+p(p+1)/2$, these bounds can be quite far apart.

There are available two methods of obtaining tests with completely known asymptotic critical points in the common case of equiprobable cells, neither of which will be pursued in detail here owing to the satisfactory result in Theorem 1. First, the methods of Dahiya and Gurland (1972) or Moore (1971) can be used to compute the exact critical points of the distribution (2.7) with λ given by (2.8). Second, the Pearson sum of squares can be replaced by the Rao-Robson statistic

$$R_n = V_n'V_n + V_n' B_n[J_n-B_n' B_n]^{-1}B_n'V_n \qquad (2.9)$$

where $J_n=J(\hat{\theta}_n)$ and $B_n=B(\hat{\theta}_n,\hat{\theta}_n)$. Rao and Robson (1974) state that the statistic (2.9) has the $\chi^2(M-1)$ asymptotic null distribution. The general proof of this assertion in Moore (1977) requires only that $r(I-qq'-BJ^{-1}B')=M-1$ and so covers the present case. The statistic R_n

is affine invariant, and in the univariate cases simulated by Rao and Robson has higher power than either $X^2(\hat{\theta}_n)$ or the Pearson-Fisher statistic.

The Rao-Robson statistic is the sum of the Pearson statistic $X^2(\hat{\theta}_n)$ and a correction term, the second term in (2.9). The correction term can be obtained as follows when $\hat{p}_{in} \equiv 1/M$ in the present case. From $\Sigma_{i=1}^{M} \partial p_i/\partial\theta_j = 0$ it follows that

$$V_n' B_n = n^{-\frac{1}{2}} \left(\sum_{i=1}^{M} \frac{N_{in}}{\hat{p}_{in}} \frac{\partial p_i}{\partial \theta_1}\bigg|_{\hat{\theta}_n}, \ldots, \sum_{i=1}^{M} \frac{N_{in}}{\hat{p}_{in}} \frac{\partial p_i}{\partial \theta_m}\bigg|_{\hat{\theta}_n} \right) \qquad (2.10)$$

and therefore when $\hat{p}_{in} \equiv 1/M$ Lemma 1 implies that

$$V_n' B_n = M n^{-\frac{1}{2}} \sum_{i=1}^{M} N_{in} d_i (0, \ldots, 0, s^{(1)}, \ldots s^{(p(p+1)/2)})$$

where $s^{(k)}$ are the entries of S^{-1} indexed in some order. Next,

$$B_n' B_n = M \sum_{i=1}^{M} d_i^2 \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & [s^{(k)} s^{(j)}] \end{array} \right]$$

where $[s^{(k)} s^{(j)}]$ is the $p(p+1)/2 \times p(p+1)/2$ matrix with the indicated products as entries. If $s = (s^{(1)}, \ldots, s^{(p(p+1)/2)})'$ and $C = Q^{-1} - M\Sigma d_i^2 [s^{(k)} s^{(j)}]$, the second term in (2.9) is $M^2 n^{-1} (\Sigma N_{in} d_i)^2 s' C^{-1} s$. The Rao-Robson statistic is computationally complex and therefore is not employed in analyzing the large data set of Section 3.

In the special case of testing bivariate circular normality, however, R_n can be easily evaluated with a programmable calculator and is worth considering for its potential additional power. The hypothesized family of density functions is now

$$f(x,y|\theta) = (2\pi\sigma^2)^{-1} e^{-\frac{1}{2\sigma^2}\{(x-\mu_1)^2 + (y-\mu_2)^2\}} \qquad (2.11)$$

parametrized by $\theta = (\mu_1, \mu_2, \sigma^2)$. The MLE's from a random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$ are $\overline{X}$, $\overline{Y}$ and

$$s^2 = \frac{1}{2n}\{\sum_{j=1}^{n}(X_j-\overline{X})^2 + \sum_{j=1}^{n}(Y_j-\overline{Y})^2\}.$$

The cells are concentric annuli centered at $(\overline{X},\overline{Y})$ with radii $c_i^{1/2}s$, where $\hat{p}_{in} \equiv 1/M$ if

$$c_i = -2\log(1-\frac{i}{M}) \qquad i=1,\ldots,M-1$$

The derivatives of p_i required for B_n are not quite as given in Lemma 1 because of the reduced number of parameters. They are

$$\frac{\partial p_{in}}{\partial \mu_1}\Bigg|_{\hat{\theta}_n} = \frac{\partial p_{in}}{\partial \mu_2}\Bigg|_{\hat{\theta}_n} = 0$$

$$\frac{\partial p_{in}}{\partial \sigma}\Bigg|_{\hat{\theta}_n} = v_i/s$$

$$v_i = 2\{(1-\frac{i}{M})\log(1-\frac{i}{M}) - (1-\frac{i-1}{M})\log(1-\frac{i-1}{M})\}.$$

Note that $v_i = 4d_i$, where d_i is as in Lemma 1 for p=2. Hence

$$B_n'B_n = \frac{M}{s^2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Sigma_1^M v_i^2 \end{bmatrix}$$

The Fisher information matrix for the family (2.11) is also diagonal,

$$J(\theta) = \sigma^{-2}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

The form of the Rao-Robson statistic stated in the following theorem is now immediate from (2.9) and (2.10).

Theorem 2: When (X_i, Y_i) have density function (2.11), the statistic

$$R_n = \frac{M}{n} \sum_{i=1}^{M} (N_{in} - \frac{n}{M})^2 + \frac{1}{n} \frac{(\Sigma_1^M N_{in} v_i)^2}{4 - M\Sigma_1^M v_i^2}$$

has the $\chi^2(M-1)$ asymptotic distribution.

3. APPLICATION TO COMMON STOCK PRICES

Suppose that P_t is the price of a security at time t. We are concerned with the variable $R_t = \log P_t - \log P_{t-1}$, which we call the return for period t. The Bachelier-Osborne model (Osborne 1959) of security price changes assumes that $\log P_t$ is a Brownian motion process, so that R_t, t=1,2,... are independent normal random variables. This model has been exhaustively examined and criticized. Many authors have suggested nonnormal distributions, especially the t-distributions and nonnormal symmetric stable distributions, as alternative models which better reflect the broad shoulders and heavy tails observed in empirical studies of returns on common stocks. See for example Blattberg and Gonedes (1974), Praetz (1969, 1972) and Fama and Roll (1968, 1971). It is common in developing theories of capital asset pricing and portfolio performance to make the additional assumption of joint normality of $(R_{1t}, \ldots, R_{kt})$ where R_{it} is the return during period t of security i. The advantage in mathematical convenience enjoyed by normal over nonnormal models is greater in multivariate settings. Multivariate normality is assumed in most presentations of models for portfolio selection (even when not necessarily essential to the general form of the model) and is essential for most distributional calculations.

See for example Fama (1976), Fama and Miller (1972), Lintner (1965), Sharpe (1970), and Tobin (1958).

Previous empirical studies of the distribution of common stock prices have been essentially univariate in nature. Fama (1976, pp. 21-35) reviews some of this work and concludes that normality is a "good working approximation" for monthly (but not daily) returns. The statistic $X^2(\hat{\theta}_n)$ of Section 2 can contribute to a multivariate analysis, as it is designed to detect multivariate versions of the types of departures from normality observed in univariate studies of stock prices. The analysis here is of course not exhaustive, both because only a single statistical procedure is applied and because there are many alternative forms of R_t which may differ in degree of nonnormality. We have made the following choices. (1) Weekly returns are used. This period is intermediate between those studied by Fama, whose data show better fit to (univariate) normality for monthly than for daily returns. Others (Praetz 1972) have suggested periods of equal market volume rather than equal duration. (2) Normality of returns R_t is tested, rather than of the relative prices $(P_t-P_{t-1})/P_{t-1}$ discussed by Fama. Experience suggests that R_t is somewhat more symmetric, but that the two variables do not differ markedly in normality. Some empirical evidence for this assertion is presented at the end of this paper. (3) The price P_t is adjusted for stock splits, but not for dividends. Financial theory includes dividends in the return, but this adjustment appears to have little impact on normality. Different choices in these three areas might lead to somewhat different conclusions.

We are grateful to Dunn and Hargitt, Inc., a Lafayette, Indiana investment advisory firm, for access to a large proprietary data base of common stock price histories. Our analysis utilizes the 560 New York Stock Exchange listings for which this data base contained adjusted midweek closing prices P_t for the 501 weeks ending April 19, 1978. With few exceptions, these are the stocks contained in the Standard & Poor 500 and in the Dow-Jones market indices. We thus have 500 observations on the 560-dimensional distribution of returns on these stocks. To assess multivariate normality, the stocks were divided at random into 140 disjoint groups of four stocks each. The test based on $X^2(\hat{\theta}_n)$ was applied within each of these groups to the quadrivariate, the four trivariate, and the six bivariate joint distributions.

These tests were conducted for both M=50 and M=25 cells, chosen to be equiprobable under the null hypothesis with estimated parameter values. These values of M are approximately the Mann-Wald (1942) recommendation for n=500 and α=0.05, and half that value. Such a choice of cells can be clearly justified in the case of testing fit to a completely specified distribution, with the smaller number of cells having generally greater power. Since data-dependent cells can be chosen equiprobable under the estimated parameter values, following the same recommendations avoids arbitrariness in cell selection. A fuller discussion of the problem of choosing cells appears in Moore (1979). Note that the use of α=0.05 in the Mann-Wald recipe is not an endorsement of that significance level for testing fit. Since the Mann-Wald recommendation decreases with α, but overstates the optimum M, the value for α=0.05 is an approximate guide for larger α as well.

For comparison, the tests for M=50 were also applied to a set of data having the same configuration as the data base, but consisting of normal deviates generated by the IMSL subroutine GGNOR. Finally, the 60 monthly returns for Xerox and IBM appearing on pp. 102 and 120 of Fama (1976) were tested for bivariate normality using M=12. These data and the cells computed from them appear in the figure. Here $X^2(\hat{\theta}_n)=13.2$, for which the P-value falls between 0.21 (from $\chi^2(10)$) and 0.28 (from $\chi^2(11)$).

For the large sample size (n=500) available for weekly returns, even distributions quite close to normal will attain a small P-value. A measure of the size of the departure from normality is therefore needed. Suppose that the true distribution of the observations X_j is F*, with mean μ^* and covariance matrix Σ^*. If θ^* corresponds to (μ^*,Σ^*), then under F* the limiting fixed cells are $E_i(\theta^*)$ (recall (2.2)), with cell probabilities

$$p_i^* = \int_{E_i(\theta^*)} dF^*.$$

It is easy to see that under F*

$$X^2(\hat{\theta}_n)/n \to \sum_{i=1}^{M} \frac{[p_i^*-p_i(\theta^*,\theta^*)]^2}{p_i(\theta^*,\theta^*)} \quad (P) \qquad (3.1)$$

where $p_i(\theta^*,\theta^*)$ is the normal cell probability computed from (2.5). (See e.g. Bishop, Fienberg and Holland 1975, p. 330 for similar statements.) Under the null hypothesis, $X^2(\hat{\theta}_n)/n \to 0(P)$. The right side of (3.1) is a measure of the closeness of F* to the normal family. It depends on the choice of cells, but approaches an integral measure of closeness as the cells are refined. Hence $X^2(\hat{\theta}_n)/n$ is an empirical measure of the magnitude of the type of departure from normality detected by the cells E_{in}. We will call $X^2(\hat{\theta}_n)/n$ the shape effect since the cells are sensitive to shape, not asymmetry. The shape effect can be restated in terms of the sample size required to declare the effect significant at a stated level, and is therefore easy to interpret. For example, Fama's data in the figure have $X^2(\hat{\theta}_n)/n=13.2/60=0.22$. Using the lower bound $\chi^2(10)$ for the critical points of $X^2(\hat{\theta}_n)$, at least 83 observations would be required to declare this effect significant at the $\alpha=0.05$ level.

Computations were carried out on a Control Data 6500. Routines in the IMSL library were employed to compute (for $p > 2$) the i/M-points c_i of the $\chi^2(p)$ distribution, and to obtain characteristic roots and vectors of the sample covariance matrix S. From these the symmetric positive definite version of $S^{-\frac{1}{2}}$ was computed and each p-variate data point X_j was transformed to $Z_j=S^{-\frac{1}{2}}(X_j-\overline{X})$. Then $Z_j'Z_j$ was computed and compared with the c_i to obtain cell counts. The Pearson statistic $X^2(\hat{\theta}_n)$ and the shape effect $X^2(\hat{\theta}_n)/n$ were obtained from the cell counts. Another IMSL rountine gave the P-values for $X^2(\hat{\theta}_n)$ from both $\chi^2(M-2)$ and $\chi^2(M-1)$.

The computer program also included a check for the lack of symmetry to which $X^2(\hat{\theta}_n)$ is insensitive. By checking the sign of each component, the number of Z_j's falling in each of the 2^p p-dimensional quadrants was determined. These quadrants are equiprobable under the estimated parameter values, and the Pearson statistic divided by n for these 2^p cells is a measure of asymmetry. This choice of cells does not share the property that narrow bounds on the limiting null distribution of the Pearson statistic are available. (The orientation of the Z_j relative to the coordinate axes, and thus the specific measure of asymmetry obtained, is dependent on the specific choice of $S^{-\frac{1}{2}}$.)

Summary statistics for 560 common stocks in 140 groups of four appear in Table 1. Although the detailed printouts show wide variations in statistical significance among the 140 groups, on the average $X^2(\hat{\theta}_n)$ is highly significant. For M=50, the $\chi^2(49)$ P-values of the mean values of $X^2(\hat{\theta}_n)$ reported in Table 1 are 0.10×10^{-3} for p=2, 0.11×10^{-6} for p=3, and 0.03×10^{-9} for p=4. The P-values for M=25 are uniformly somewhat smaller in the individual tests, as well as for the mean Pearson statistic. This conforms to the expectation that an M half that recommended by Mann and Wald somewhat improves the sensitivity of the test. In contrast, the P-value of the mean of $X^2(\hat{\theta}_n)$ for the simulated data is close to 0.5, and examination of the individual tests confirms that the χ^2 distribution fits closely.

Low P-values do not imply large deviations from normality when n=500. The mean size of the shape effect $X^2(\hat{\theta}_n)/n$ for the common stocks as a multiple of that for the simulated data is 2 for p=2, 2.5 for p=3 and 3 for p=4 in the M=50 case. That is, whereas about 675 observations would be needed to declare the mean effect for the simulated data significant at α=0.05, the corresponding sample sizes for the stock data are at least 345 for p=2, 275 for p=3, and 225 for p=4. This compares favorably with the results for the smaller sample of monthly relative prices from Fama. Indeed, if the degree of asymmetry, is similarly measured by the ratio of the mean of the asymmetry effect for the stock return data to that for the simulated data, the asymmetry effect for p=2 is as great as the shape effect. As p increases, the degree of asymmetry decreases, while nonnormality in shape increases. We do not take this as an indication of serious asymmetry, but as suggesting that the nonnormality in tails and shoulders which has led to the suggestion that other symmetric laws be adopted is no more serious than is violation of the symmetry assumption common to the competing models. [Note that the sizes of the shape and asymmetry effects cannot be directly compared, as they refer to different configurations of cells. This fact motivates the procedure of comparing both to the values they took in the simulated data.]

In summary, the multivariate analysis carried out here seems to sustain Fama's conclusion that normality of returns on common stocks is a good working approximation. (Of course the null hypothesis of exact normality can be rejected at all common significance levels for most-but not all-of the groups of stocks examined.) The summary statistics in Table 1 allow others to make further analyses. Two facts not apparent from Table 1 should be mentioned. First, the cell counts confirm that the nature of the deviation from normality is broad, rather flat, shoulders and heavy tails. Cell counts other than for cells near the sample mean and the last cell in the tails generally show only random deviations from the expected 10 (M=50) or 20 (M=25) observations per cell. Second, while the mean shape effect increases with the dimension of the joint distribution under study, radical nonnormality for p=4 can almost always be attributed to radical nonnormality in a univariate or bivariate marginal distribution. Thus extending our study to $p > 4$ did not seem justified. We investigated the relatively few stocks with strikingly nonnormal returns, but found no common feature (such as low dollar value of outstanding stock) which might account for lack of normality.

To what extent can the conclusion that multivariate normality is a good working approximation for common stock returns be generalized? Fama's discussion suggests that returns for longer periods will be no less normal than weekly returns. The effect of redefining P_t to include dividends is not known. To gain insight into the effect of testing the relative price $(P_t - P_{t-1})/P_{t-1}$ rather than R_t, we analyzed the relative prices for the first 20 of the 140 randomly chosen groups of four stocks each. Table 2 presents summary statistics for both returns and relative prices. The sharply greater asymmetry of the relative price data is apparent, especially for p=2. In shape, however, the two variables are quite similar.

APPENDIX

Proof of Lemma 1: The differentiation of $p_i(\theta,\theta_0)$ is most easily carried out in matrix form, making free use of the tables of formulae given by Dwyer (1967). Applying the transformation $z=\Sigma_0^{-\frac{1}{2}}(x-\mu_0)$ to (2.5),

$$p_i(\theta,\theta_0) = (2\pi)^{-p/2} \frac{|\Sigma_0|^{\frac{1}{2}}}{|\Sigma|^{\frac{1}{2}}} \int_{c_{i-1}\le z'z<c_i} e^{-\frac{1}{2}(\Sigma_0^{\frac{1}{2}}z+\mu_0-\mu)'\Sigma^{-1}(\Sigma_0^{\frac{1}{2}}z+\mu_0-\mu)} dz.$$

By Dwyer's (11.1) applied under the integral sign,

$$\left.\frac{\partial p_i(\theta,\theta_0)}{\partial\mu}\right|_{\theta=\theta_0} = (2\pi)^{-p/2}\, \Sigma_0^{-\frac{1}{2}} \int_{c_{i-1}\le z'z<c_i} e^{-\frac{1}{2}z'z}\, dz .$$

and the integral (a p-vector) is 0 by symmetry.

To compute the matrix of derivatives $\partial p_i/\partial\Sigma$, again differentiate under the integral sign. Writing $u=x-\mu$ in (2.5), the required derivative is

$$\frac{\partial}{\partial\Sigma}\{|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}u'\Sigma^{-1}u}\} = \tfrac{1}{2}|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}u'\Sigma^{-1}u}\{\Sigma^{-1}uu'\Sigma^{-1}-\Sigma^{-1}\}$$

by the product rule and (11.3), (11.8) of Dwyer (1967). Therefore

$$\left.\frac{\partial p_i}{\partial\Sigma}\right|_{\theta=\theta_0} = (2\pi)^{-p/2}\int_{E_i(\theta_0)}\frac{\partial}{\partial\Sigma}\{|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}u'\Sigma^{-1}u}\}\Big|_{\theta=\theta_0} du$$

$$= \tfrac{1}{2}\Sigma_0^{-1}\{(2\pi)^{-p/2}|\Sigma_0|^{-\frac{1}{2}}\int_{E_i(\theta_0)} uu'\, e^{-\frac{1}{2}u'\Sigma^{-1}u}du\}\Sigma_0^{-1}$$

$$-\tfrac{1}{2}p_i(\theta_0,\theta_0)\Sigma_0^{-1}$$

Now setting $u=\Sigma_0^{\frac{1}{2}}z$ in the first expression yields

$$\left.\frac{\partial p_i}{\partial\Sigma}\right|_{\theta=\theta_0} = \tfrac{1}{2}\Sigma_0^{-\frac{1}{2}}\{(2\pi)^{-p/2}\int_{c_{i-1}\le z'z<c_i} zz'\, e^{-\frac{1}{2}z'z}\, dz\}\Sigma_0^{-\frac{1}{2}}$$

$$-\tfrac{1}{2}p_i(\theta_0,\theta_0)\Sigma_0^{-1}$$

The $p \times p$ matrix within the brackets has the form $a_i I_p$ for some number a_i. Therefore

$$\left.\frac{\partial p_i}{\partial\Sigma}\right|_{\theta=\theta_0} = \tfrac{1}{2}(a_i-p_i(\theta_0,\theta_0))\Sigma_0^{-1} \tag{A.1}$$

and it remains only to evaluate a_i. Introduce p-dimensional spherical coordinates $0 \le r < \infty$, $0 \le \theta < 2\pi$, $0 \le \varphi_1,\ldots,\varphi_{p-2} < \pi$. Then $z_p = r\cos\varphi_{p-2}$ and

$$dz = r^{p-1}\left(\prod_{j=1}^{p-2}\sin^j\varphi_j\right)dr\, d\theta\, d\varphi_1\ldots d\varphi_{p-2}.$$

Therefore

$$a_i = (2\pi)^{-p/2} \int_{c_{i-1} \le z'z < c_i} z_p^2 \, e^{-\frac{1}{2}z'z} \, dz$$

$$= 2(2\pi)^{-p/2}\left\{\int_{c_{i-1}^{\frac{1}{2}}}^{c_i^{\frac{1}{2}}} r^{p+1} \, e^{-\frac{1}{2}r^2} \, dr\right\} A_{p-2} \prod_{j=0}^{p-3} B_j \qquad (A.2)$$

where

$$A_{p-2} = \int_0^\pi \cos^2 \varphi \, \sin^{p-2} \varphi \, d\varphi$$

$$B_j = \int_0^\pi \sin^j \varphi \, d\varphi$$

(and note that $2B_0=\int_0^{2\pi} d\theta$). Now $A_{p-2}=B_{p-2}/p$ and

$$\int r^{p+1} \, e^{-\frac{1}{2}r^2} \, dr = -r^p \, e^{-\frac{1}{2}r^2} + p \int r^{p-1} \, e^{-\frac{1}{2}r^2} \, dr$$

and

$$p_i(\theta_0, \theta_0) = (2\pi)^{-p/2} \int_{c_{i-1} \le z'z < c_i} e^{-\frac{1}{2}z'z} \, dz$$

$$= 2(2\pi)^{-p/2}\left\{\int_{c_{i-1}^{\frac{1}{2}}}^{c_i^{\frac{1}{2}}} r^{p-1} \, e^{-\frac{1}{2}r^2} \, dr\right\} \prod_{j=0}^{p-2} B_j .$$

Substituting these relations into (A.2) gives

$$a_i = 2(2\pi)^{-p/2} \, p^{-1} \prod_{j=0}^{p-2} B_j \{c_{i-1}^{p/2} \, e^{-c_{i-1}/2} - c_i^{p/2} \, e^{-c_i/2}\} + p_i(\theta_0, \theta_0)$$

Finally, $B_0=\pi$, $B_1=2$ and $B_j=(j-1)B_{j-2}/j$ imply that

$$\prod_{j=0}^{p-2} B_j = (2\pi)^{(p-2)/2} \pi [(p-2)(p-4)\ldots(4)(2)]^{-1} \qquad p \text{ even}$$

$$= (2\pi)^{(p-1)/2} [(p-2)(p-4)\ldots(5)(3)]^{-1} \qquad p \text{ odd.}$$

Thus the number $d_i=(a_i-p_i(\theta_0,\theta_0))/2$ in (A.1) has the value stated in Lemma 1.

Proof that the nonzero root of $BJ^{-1}B'$ is $2Mp\Sigma_1^M d_i^2$ when c_i are the i/M points of $\chi^2(p)$: Because this root does not depend on μ_0 and Σ_0, we can assume that $\mu_0=0$, $\Sigma_0=I_p$ and cells are bounded by $x'x=c_i$. Let us order the components of θ as follows. First the p means $\mu_1,\ldots,\mu_p$, then $\sigma_{11},\ldots,\sigma_{pp}$ and then the p(p-1)/2 upper triangular off-diagonal elements of Σ in an arbitrary order. Then (2.4) and the reference cited there show that for $\mu_0=0$, $\Sigma_0=I_p$ the matrix J^{-1} is diagonal with the first p entries 1, the succeeding p entries 2, and the remaining p(p-1)/2 entries again 1. The matrix B has p zero columns, then p columns each equal to $M^{\frac{1}{2}}(d_1,\ldots,d_M)'$, then p(p-1)/2 zero columns. Therefore

$$BJ^{-1}B' = 2Mp[d_i d_j]_{M\times M}$$

The matrix of rank 1 $[d_i d_j]$ has characteristic equation $(-1)^M\delta^{M-1}(\delta-a)=0$. That $a=\Sigma_1^M d_i^2$ will be shown by induction. The result is immediate in the 2×2 case. Suppose it holds for $(M-1)\times(M-1)$. Expanding $D_M=|[d_i d_j]-\delta I|$ in cofactors of the last column,

$$D_M = (d_M^2-\delta)D_{M-1} + g(\delta)$$

where the linear combination of cofactors $g(\delta)$ can be seen to be a polynomial of degree M-2 in δ. The characteristic equation is therefore by induction,

$$(d_M^2-\delta)(-1)^{M-1}\delta^{M-2}(\delta-\sum_{i=1}^{M-1} d_i^2) + g(\delta) = 0$$

Since this equation is known to have the form $(-1)^M\delta^{M-1}(\delta-a)=0$, all terms of degree less than M-1 must cancel, and can be ignored. The terms of degree M and M-1 are $(-1)^M\delta^M + (-1)^{M-1}\delta^{M-1}\Sigma_1^M d_i^2$, so that $a=\Sigma_1^M d_i^2$ follows.

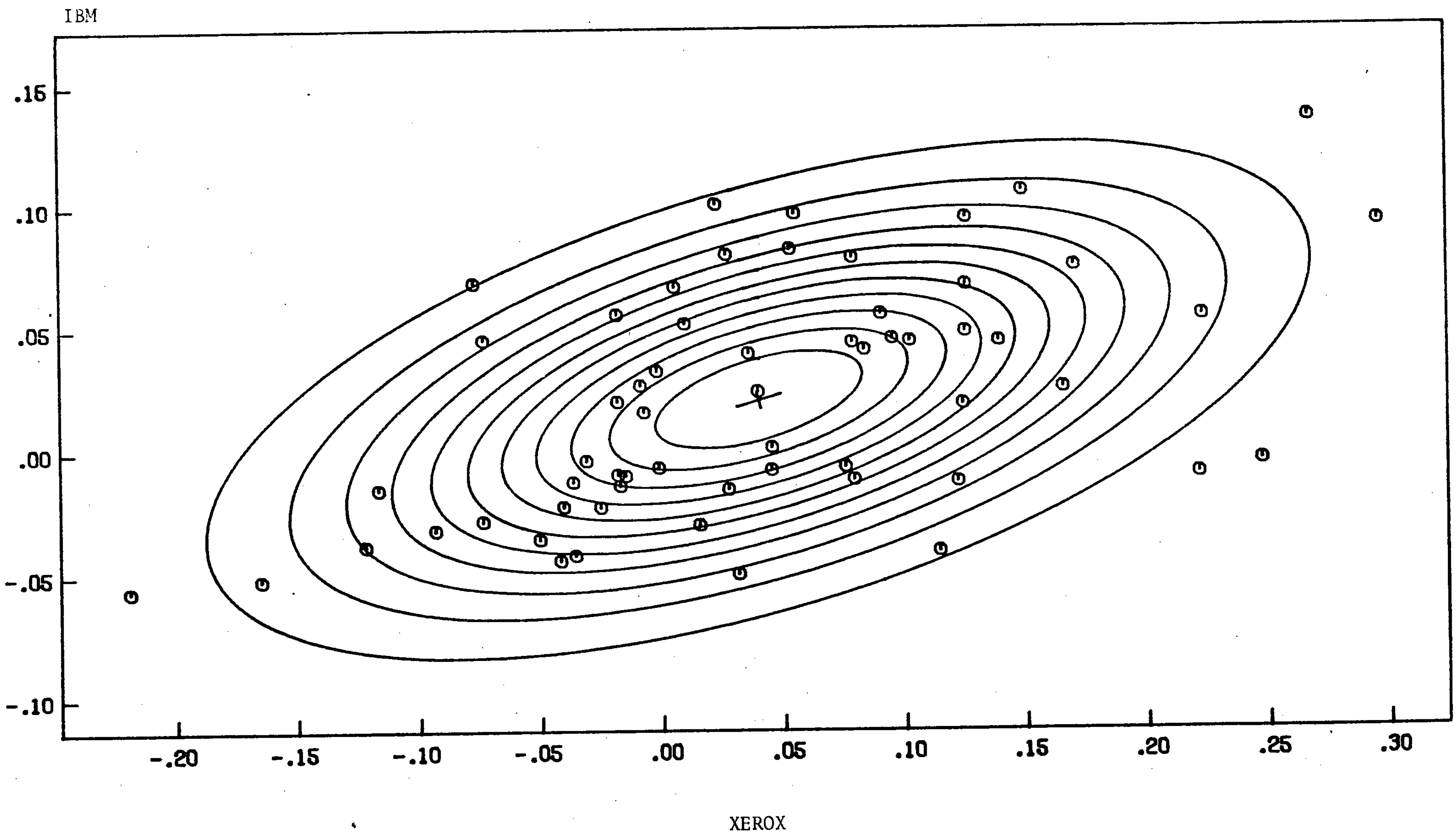
1. Summary Statistics, Stock Returns and Simulation

Statistic	Simulated M=50	Stock returns M=50	Stock returns M=25
		p=2	
Mean Pearson	48.070	94.524	63.730
St. Dev. Pearson	9.911	27.862	24.136
Min Pearson	23.600	37.000	14.700
Max Pearson	86.000	258.200	240.500
Mean shape effect	.096	.189	.127
St. Dev. shape effect	.020	.056	.048
Mean asymmetry	.003	.007	.007
St. Dev. asymmetry	.002	.006	.006
Min asymmetry	.000	.000	.000
Max asymmetry	.017	.039	.039
N of observations	840.	840.	840.
		p=3	
Mean Pearson	47.696	118.452	84.516
St. Dev. Pearson	9.680	33.807	29.150
Min Pearson	22.800	41.600	25.600
Max Pearson	81.200	245.400	209.600
Mean shape effect	.095	.237	.169
St. Dev. shape effect	.019	.068	.058
Mean asymmetry	.008	.015	.015
St. Dev. asymmetry	.004	.008	.008
Min asymmetry	.001	.001	.001
Max asymmetry	.031	.057	.057
N of observations	560.	560.	560.
		p=4	
Mean Pearson	48.671	145.286	106.645
St. Dev. Pearson	10.664	39.399	34.030
Min Pearson	28.400	72.600	43.800
Max Pearson	78.600	264.400	207.900
Mean shape effect	.097	.291	.213
St. Dev. shape effect	.021	.079	.068
Mean asymmetry	.020	.030	.030
St. Dev. asymmetry	.007	.011	.011
Min asymmetry	.008	.009	.009
Max asymmetry	.037	.069	.069
N of observations	140.	140.	140.

2. Comparison of Returns and Relative Prices (M=50)

Statistic	Returns R_t	Relative Prices $(P_t - P_{t-1})/P_{t-1}$
	p=2	
Mean Pearson	97.983	98.468
St. Dev. Pearson	28.816	30.729
Min Pearson	47.000	41.600
Max Pearson	220.400	242.200
Mean shape effect	.196	.197
St. Dev. shape effect	.058	.061
Mean asymmetry	.006	.012
St. Dev. asymmetry	.005	.009
Min asymmetry	.000	.000
Max asymmetry	.023	.040
N of observations	120.	120.
	p=3	
Mean Pearson	122.307	128.325
St. Dev. Pearson	32.237	32.855
Min Pearson	58.200	64.800
Max Pearson	219.200	254.000
Mean shape effect	.245	.257
St. Dev. shape effect	.064	.066
Mean asymmetry	.015	.022
St. Dev. asymmetry	.008	.011
Min asymmetry	.001	.003
Max asymmetry	.037	.056
N of observations	80.	80.
	p=4	
Mean Pearson	155.400	152.760
St. Dev. Pearson	30.202	32.678
Min Pearson	115.200	102.800
Max Pearson	220.600	218.600
Mean shape effect	.311	.306
St. Dev. shape effect	.060	.065
Mean asymmetry	.031	.041
St. Dev. asymmetry	.010	.017
Min asymmetry	.015	.015
Max asymmetry	.053	.078
N of observations	20.	20.

Data-Dependent Cells for a Bivariate Sample
60 Monthly Relative Prices of IBM and Xerox



REFERENCES

Andrews, D. F., Gnanadesikan, R., and Warner, J. L. (1973), "Methods For Assessing Multivariate Normality," in Multivariate Analysis-III, ed. Paruchuri R. Krishnaiah, New York: Academic Press, 95-116.

Bishop, Yvonne M. M., Fienberg, Stephen E., and Holland, Paul W. (1975), Discrete Multivariate Analysis, Cambridge, Massachusetts: The MIT Press.

Blattberg, Robert C., and Gonedes, Nicholas J. (1974), "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices," Journal of Business, 47, 244-280.

Chernoff, Herman, and Lehmann, E. L. (1954), "The Use of Maximum Likelihood Estimates in χ^2 Tests for Goodness of Fit," Annals of Mathematical Statistics, 25, 579-586.

Dahiya, Ram C., and Gurland, John (1972), "Pearson Chi-Square Test of Fit with Random Intervals," Biometrika, 59, 147-153.

Dwyer, Paul S. (1967), "Some Applications of Matrix Derivatives in Multivariate Analysis," Journal of the American Statistical Association, 62, 607-625.

Fama, Eugene F. (1976), Foundations of Finance: Portfolio Decisions and Securities Prices, New York: Basic Books.

____, and Miller, Merton H. (1972), The Theory of Finance, New York: Holt, Rinehart and Winston.

____, and Roll, Richard (1968), "Some Properties of Symmetric Stable Distributions," Journal of the American Statistical Association, 63, 817-836.

____, and Roll, Richard (1971), "Parameter Estimates for Symmetric Stable Distributions," Journal of the American Statistical Association, 66, 331-338.

Lintner, John (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," The Review of Economics and Statistics, 47, 13-37.

Mann, H. B., and Wald, Abraham (1942), "On the Choice of the Number of Class Intervals in the Application of the Chi-Square Test," Annals of Mathematical Statistics, 13, 306-317.

Moore, David S. (1971), "A Chi-Square Statistic with Random Cell Boundaries," Annals of Mathematical Statistics, 42, 147-156.

_____, (1977), "Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests of Fit," Journal of the American Statistical Association, 72, 131-137.

_____ (1979), "Chi-Square Techniques," in Goodness of Fit Techniques, ed. Ralph D'Agostino and M. A. Stephens New York: Marcel Dekker, Inc., to appear.

_____, and Spruill, M. C. (1975), "Unified Large-Sample Theory of General Chi-Squared Statistics for Tests of Fit," Annals of Statistics, 3, 599-616.

Osborne, M. F. M. (1959), "Brownian Motion in the Stock Market," Operations Research, 7, 145-173.

Praetz, Peter D. (1969), "Australian Share Prices and the Random Walk Hypothesis," Australian Journal of Statistics, 11, 123-139.

_____ (1972), "The Distribution of Share Price Changes," Journal of Business, 45, 49-55.

Press, S. James (1972a), Applied Multivariate Analysis, New York: Holt, Rinehart and Winston.

_____ (1972b), "Multivariate Stable Distributions," Journal of Multivariate Analysis, 2, 444-462.

Rao, K. C., and Robson, Douglas S. (1974), "A Chi-Square Statistic for Goodness-of-Fit within the Exponential Family," Communications in Statistics, 3, 1139-1153.

Roy, A. R. (1956), "On χ^2 Statistics with Variable Intervals," Technical Report No. 1, Stanford University Department of Statistics.

Sharpe, William F. (1970), Portfolio Theory and Capital Markets, New York: McGraw-Hill.

Tobin, James (1958), "Liquidity Preference as Behavior Toward Risk," Review of Economic Studies, 26, 65-86.

Watson, Geoffry S. (1957), "The Chi-Squared Goodness-of-Fit Test for Normal Distributions," Biometrika, 44, 336-348.

_____ (1958), "On Chi-Square Goodness-of-Fit Tests for Continuous Distributions," Journal of the Royal Statistical Society, Series B, 20, 44-61.

_____ (1959), "Some Recent Results in Chi-Square Goodness-of-Fit Tests," Biometrics, 15, 440-468.