THE JOURNAL OF FINANCE • VOL. XXXIX, NO. 1 • MARCH 1984

Models of Stock Returns—A Comparison

STANLEY J. KON*

ABSTRACT

In this paper a discrete mixture of normal distributions is proposed to explain the observed significant kurtosis (fat tails) and significant positive skewness in the distribution of daily rates of returns for a sample of common stocks and indexes. Stationarity tests on the parameter estimates of this discrete mixture of normal distributions model revealed significant differences in the mean estimates that can explain the observed skewness and significant differences in the variance estimates that can explain the observed kurtosis. An alternative explanation for the observed fat tails is the symmetric student model. The result of a comparison between the models is that the discrete mixture of normal distributions model has substantially more descriptive validity than the student model.

FOR MANY YEARS both financial economists and statisticians have been concerned with the description of stock market returns. The form of the distribution of stock returns is a crucial assumption for mean-variance portfolio theory, theoretical models of capital asset prices, and the prices of contingent claims. For example, understanding the behavior of the variance is essential to option pricing models. Empirical tests of asset pricing models and the efficient markets hypothesis draw statistical inferences that are also conditional on distributional assumptions. The most convenient assumption for financial theory and empirical methods is that the distribution of security rates of return be multivariate normal with parameters that are stationary over time. Since the normal distribution is stable under addition, any arbitrary portfolio of stocks will also be normally distributed. With the additional assumption of risk aversion, mean-variance theory follows. Furthermore, the assumptions of normality and parameter stationarity are required for most of the econometric techniques typically used in empirical research.

Tests of the normality hypothesis on the *daily* returns of the Dow Jones Industrial stocks by Fama [12] revealed more kurtosis (fatter tails) than that predicted from a sample of independent and identically distributed normal variates. Fama concluded from this evidence that the distribution of price changes conforms better to the stable Paretian distribution with characteristic exponent less than 2. The evidence provided by Blattberg and Gonedes [4], however,

* University of Michigan, Graduate School of Business Administration. The author wishes to thank C. Ball, M. Bradley, S. Chaplinsky, R. Hagerman, A. Kalay, E. H. Kim, E. Melnick, G. Simon, A. Tschoegl, M. Weinstein, J. Williams, and the participants in the Finance Workshop at The University of Michigan for their helpful comments and discussions and Steven Freund for his excellent programming assistance. This paper has especially benefitted from the comments of anonymous referees. This research was partially supported by a summer research grant from the University of Michigan Graduate School of Business Administration.

indicates that the distribution of *monthly* returns conforms well to the normality hypothesis. Therefore, most of the empirical research in the finance literature proceeded to use monthly data with reasonable confidence in their inferences. More recently, researchers are using daily data in order to isolate information events (e.g., Charest [7] and Aharony and Swary [1]) or for the advantage of a large estimation sample for statistical reliability in asset pricing model tests (e.g., Roll and Ross [29]). This potential for more powerful empirical tests, however, is still conditional on the yet unresolved distribution of security returns issue.

In order to test the validity of the stable Paretian hypothesis, subsequent empirical studies concentrated on its stability property. Officer [25], Hsu et al. [17], Blattberg and Gonedes [4], and Hagerman [16] all report evidence that is *not* consistent with the stable Paretian hypothesis. The evidence indicates that the characteristic exponents of the distribution of the intertemporal sum for the returns on stocks and portfolios rise with the sum size. This result is a clear violation of the stability property.

Alternative explanations for the observed fat tails in the empirical distribution of security returns involve model specifications in which the true underlying generating process is a mixture of normals. Several researchers have postulated a continuous mixture of normal distributions where the variance is a random variable. Praetz [27] and Blattberg and Gonedes [4] prove that if the variance of the normal follows an inverted gamma distribution, then the resulting (posterior) distribution is the student. A rigorous comparison of the student and stable distributions was made by Blattberg and Gonedes [4] using daily rate of return data on each of the 30 stocks in the Dow Jones Industrial Average. The results clearly indicate that the student model has greater descriptive validity than the symmetric-stable model.[1]

Considerably less attention has been given to a discrete mixture of normal distributions even though the economic scenarios may be more reasonable. Recall that the empirical evidence on the distribution of daily stock returns clearly rejects the *stationary normal* distribution model. It is, however, the normality hypothesis that is crucial to models of financial theory. Stationarity is merely a convenient sampling assumption. In fact, theory predicts that changes in the investment and financial decision variables of firm managers will result in adjustments to the expected return and standard deviation parameters of the distribution of a security's return. Boness et al. [5] found that with weekly return data before and after a capital structure change, the parameters of the price change processes do indeed shift. They also present evidence that the overall series contain substantially more departures from normality than the series of either the pre or post capital structure change subperiods. Furthermore, Christie [10] demonstrates that the standard deviation of a stock's return is an increasing function of financial and operating leverage and empirically verifies the financial leverage effect.

[1] Other models of stock returns include the compound events model proposed by Press [28] and the lognormal-normal model proposed by Clark [8]. However, the probability densities in neither of these models have exact solutions, and hence, they will not be considered as candiate models for the statistical methodology in this paper.

There are also information signal scenarios concerning the disclosure of a firm's quarterly earnings that lead to parameter shifts. These seasonal announcements result in rate of return observations with higher variance during the disclosure period than during the nonannouncement periods. This is empirically verified by Beaver [3] and others with realized stock returns and in the market's ex ante assessments of the variance around these announcements by Patell and Wolfson [26]. Since this information signal of parameter shifts can be generalized to all firm-specific events, Christie [9] has formulated a discrete mixture of two normal distributions model where returns drawn from the distribution with the higher variance represent information events while the other distribution generates noninformation random variables. This model was successfully applied to information announcements in the *Wall Street Journal*. Ball and Torous [2] also derive and provide evidence consistent with a mixture of two normal distributions model for daily returns resulting from a Bernoulli jump process to describe information arrivals.

The generating process of security returns is further complicated by exogenous macro information and institutional trading restrictions. For example, using monthly return data, both Officer [25] and Hsu et al. [17] provide evidence that a substantial increase in the characteristic exponent of common stocks occurs from the pre World War II period to the postwar period. More recently, French [13] and Gibbons and Hess [14] have found a significant difference in the mean return of Mondays compared to the other days of the week. These authors and Fama [12] also find that the standard deviation of Monday returns is higher than those of the other days of the week. This latter result is supported by the intuition that more information relevant to price formation will be accumulated over the weekend than just overnight for the rest of the week's trading days. Furthermore, Keim [18] documents nonstationary mean excess returns related to January and the first trading week of that month. These macro components may result in a mixture of normals for the market portfolio and a mixture of more than two normals for the total return distribution on individual stocks. For example, returns may be drawing from a noninformation distribution, a firm-specific information distribution, and a market-wide information distribution—hence, a mixture of three normal distributions. The actual number of normal distributions is itself an empirical issue and may vary across firms.

The purpose of this paper is to provide evidence on the descriptive validity of a discrete mixture of normal distributions process as a statistical model for stock returns. This is accomplished by estimating the parameters of the respective models for mixtures of $N = 1, 2, 3, 4$, and 5 normal distributions. Likelihood ratio tests of model specification indicate that the sample of 30 Dow Jones stocks can be described by a mixture of four normal distributions for 7 stocks, a mixture of three models for 11 stocks, and a mixture of two normal distributions for the remaining 12 stocks. This model is also compared with the currently most empirically descriptive alternative explanation for the observed fat tails in stock returns—the symmetric student distribution model (see Blattberg and Gonedes [4]). The log-odds results indicate that the discrete mixture of normal distributions model has substantially more descriptive validity than the student model.

The remainder of this paper is organized as follows. In Section I, we present some empirical evidence on the returns of stock indexes that demonstrates the potential for the discrete mixture of normal distributions hypothesis to explain their observed fat tails. Section II contains a discussion of model specification and parameter estimation for the generalized discrete mixture of normal distributions. This section also describes the statistical methods for inferring the number of normals generating the data. Section III presents the empirical evidence and Section IV contains a summary of the results and a discussion of the implications for empirical research in financial economics.

I. Preliminary Evidence on the Discrete Mixture of Normal Distributions Hypotheses

In the previous section, it was argued that the true distribution of stock returns may be normal, but that its parameters shift among a finite set of values. There are *time-ordered* shifts associated with capital structure changes, acquisitions, stock splits, or exogenous market events; and *cyclical* shifts between sets of parameters, as in the day of the week effect or the seasonal announcements of firm earnings and dividends. In order to assess the potential impact of both types of shifts on the distributional issue, an 18½-year time-series of 4,639 daily return observations on the Standard and Poor's Composite (S&P), the CRSP value-weighted (VW), and the CRSP equal-weighted (EW) indexes are partitioned (1) by year to roughly account for time-ordered events, (2) by day of the week to account for cyclical events, and (3) by year and by day of the week to account for both effects.

All three indexes exhibited significant skewness and kurtosis (fat tails) statistics at the 0.01 probability level for the entire sample period (see Table I). The observed skewness may be explained by shifts in the mean parameter in the time-series and the observed fat tails are consistent with shifts in the variance parameter (see Christie [9], Appendix A). Partitioning the data into annual subperiods reduced the frequency of rejecting the stationary normality null hypothesis. The partition by day of the week still rejected the null hypothesis, but with somewhat less significance than for the entire sample period.[2]

Since both of the previous partitions led to a reduction in the kurtosis coefficient test statistics, it is worth trying a partition of the entire sample period by year *and* by day of the week. For a sample of this size (about 50 observations) a studentized range value greater than 5.77 should occur only once in every 100 repetitions.[3] The 18 years of S&P returns contain 90 subsets of data. In only 8 subsets could the stationary normal null hypothesis be rejected. For the VW

[2] The individual statistics are not reported for brevity, but are available from the author upon request.

[3] Since the position of the percentage points for the kurtosis coefficient in samples of less than 200 has not been established, the studentized range test is preferable for detecting fat tails. The studentized range is never more effective than the kurtosis coefficient in detecting departures from normality in the tails of a distribution for sample sizes greater than 200. See David et al. [11]. Therefore, throughout this paper, whenever the sample size is greater than 200, inferences will be made with the kurtosis coefficient. Only in the by year and by day of the week case does the sample size fall below 200, necessitating the use of the studentized range statistic.

Table I

Tests for Departure from Normality

I.D.	Security (or index)	Mean	Standard Deviation	Skewness Statistic[a]	Kurtosis Coefficient[b]	Number of Observations
1	Allied Chemical Corp	0.0406	1.7066	0.4181	6.6345	4639
2	Aluminum Co America	0.0362	1.6394	−0.0459	6.3552	4639
3	American Brands Inc	0.0509	1.2903	0.5232	7.2854	4639
4	American Can Co	0.0223	1.1590	0.1796	6.3436	4639
5	American Tel & Teleg	0.0254	0.9075	0.5139	7.8777	4637
6	Bethelem Steel Corp	0.0301	1.6565	0.0744	9.0226	4639
7	Du Pont	0.0235	1.2953	0.3278	5.6329	4637
8	Eastman Kodak Co	0.0459	1.5067	0.3255	6.2847	4639
9	Exxon Corp	0.0534	1.1339	0.1860	4.8022	4639
10	General Electric Co	0.0370	1.3767	0.1679	5.6936	4639
11	General Foods Corp	0.0242	1.3676	0.4098	6.3881	4639
12	General Motors Corp	0.0299	1.2844	0.2397	5.9173	4637
13	Goodyear	0.0294	1.5503	0.2479	5.0644	4639
14	INCO Ltd	0.0253	1.5662	0.1538	5.6975	4639
15	Inter Business Mach	0.0430	1.3261	0.3348	5.6165	4635
16	Inter Harvester Co	0.0350	1.5223	0.0678	5.7235	4639
17	Inter Paper Co	0.0407	1.6089	0.2983	5.5149	4639
18	Johns Manville Corp	0.0352	1.6897	0.3240	6.4031	4639
19	Merck & Co Inc	0.0626	1.4253	0.2868	5.6924	4639
20	Minnesota Mng & Mfg	0.0378	1.3900	0.3140	5.9065	4639
21	Owens Illinois Inc	0.0334	1.5468	0.1764	6.5306	4639
22	Proctor & Gamble Co	0.0336	1.1536	0.3799	7.1343	4639
23	Sears Roebuck & Co	0.0188	1.3089	0.3302	6.6317	4639
24	Standard Oil Co Cal	0.0612	1.4062	0.3300	6.0832	4639
25	Texaco Inc	0.0463	1.4173	0.2809	5.9354	4639
26	Union Carbide Corp	0.0305	1.3917	0.3415	5.9528	4639
27	United Aircraft Prod	0.1042	2.8874	0.8791	6.8732	4629
28	United Sts Stl Corp	0.0298	1.5819	0.9080	11.5863	4639
29	Westinghouse Elc Co	0.0506	1.9172	−0.0721	13.9385	4639
30	Woolworth F W Co	0.0359	1.6792	0.7257	10.5653	4639
31	Standard & Poor's 500	0.0221	0.7702	0.2617	5.8410	4639
32	Value Weighted Mkt	0.0387	0.7595	0.1434	6.0603	4639
33	Equal Weighted Mkt	0.0741	0.7871	−0.1016	8.8296	4639

[a] Skewness statistic is the third central moment divided by the three-halves power of the second central moment. The upper and lower one percentage points of its distribution are 0.084 and −0.084, respectively.

[b] The kurtosis coefficient is the fourth central moment divided by the square of the second central moment. The upper and lower one percentage points of its distribution are 3.18 and 2.14, respectively.

index, there were only 9 rejections of the null hypothesis. In the 90 subsets of the EW returns, there were 26 rejections of normality. The larger frequency of rejection for the EW returns may indicate that smaller firms have fewer but more surprising information releases.

The tests of normality on the partition of the data by day of the week *and* by year provide a strong motivation for pursuing the discrete mixture of normal distributions hypothesis. Since there is no reason to believe that time-ordered shifts only take place on January 1, this partition is certainly not the true data generating model. But it does indicate that a model specification intended to represent the true mixture process must be able to accommodate both cyclical and structural (time-ordered) shifts in the two parameters of a normal distribution. The model specification and estimation procedures described in the next section will satisfy these requirements.

II. A General Model for a Discrete Mixture of Normal Distributions

Given the empirical evidence of the previous section, the formulation of a discrete mixture of normal distributions model must allow discontinuous shifts in the true parameter values to successively new levels (structural shifts) as well as shifts back and forth between a particular set of parameter values (cyclical shifts). Alternatively stated, each return observation is a drawing from one of N sets of parameter values. As long as we refer to subsets of the data whose observations are not necessarily consecutive in time, then both the structural and cyclical type parameter shifts can be accommodated.

The generalized discrete mixture of normal distributions model views each return observation on a stock, r_t, as having been generated by one of the following N distinct equations:

$$\begin{aligned} r_t &= \alpha_1 + u_{1t} \qquad && t \in I_1 \\ r_t &= \alpha_2 + u_{2t} && t \in I_2 \\ &\vdots && \vdots \\ r_t &= \alpha_N + u_{Nt} && t \in I_N \end{aligned} \tag{1}$$

where I_i, $i = 1, 2, \cdots, N$ are the homogeneous information sets with T_i observations in each set. $\sum_{t=1}^{N} T_i = T$, u_{it} is independently and identically normally distributed with a mean of zero of variance of σ_i^2, $0 < \sigma_i^2 < \infty$, $i = 1, 2, \cdots, N$. Define $\lambda_i = T_i/T$ as the proportion of observations associated with information set I_i. Then, for a given N, the parameter vector, $\underline{\theta} = \{\alpha_1, \alpha_2, \cdots, \alpha_N, \sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2, \lambda_1, \lambda_2, \cdots, \lambda_{N-1}\}$, can be estimated by maximizing the likelihood function

$$l(\underline{\theta}/\underline{r}) = \prod_{t=1}^{T} \left[\sum_{i=1}^{N} \lambda_i p(r_t \mid \underline{\gamma}_i)\right], \tag{2}$$

where $\underline{r} = (r_1, r_2, \cdots, r_T)'$, $\underline{\gamma}_i = (\alpha_i, \sigma_i^2)$, and $p(r_t \mid \underline{\gamma}_i)$ is a normal probability density function with mean α_i and variance σ_i^2. The details of the estimation procedure and generality of the model specification are contained in the Appendix.

The generalized discrete mixture of normal distributions model and estimation procedure has been presented for any number N of normals. Given N, the maximum likelihood procedure estimates the $3N - 1$ parameters defining that particular specification. In order to determine N empirically, the model can be estimated for each $N = 1, 2, \cdots$, and specifications can be compared via the likelihood ratio test.

Consider the likelihood function $l(\underline{\theta} \mid \underline{r})$ in Equation (2) and define $l(\underline{\theta}_i \mid \underline{r})$ as that likelihood function for the case where $N = i$. For a pairwise test between the different specifications, i and j, we can use the generalized likelihood-ratio

$$\Lambda_{ij} = \frac{l(\underline{\theta}_i \mid \underline{r})}{l(\underline{\theta}_j \mid \underline{r})}, \qquad i < j. \tag{3}$$

For example, the likelihood ratio for a test of the stationary normal model against the mixture of two normal distributions alternative is $\Lambda_{12} = l(\underline{\theta}_1 \mid \underline{r})/l(\underline{\theta}_2 \mid \underline{r})$. Note that the parameter vector for the stationary normal model is a subset of the parameter vector for the mixture of two normals model. That is, when $\lambda_1 =$ 0, 1 or $\gamma_1 = \gamma_2$, $l(\underline{\theta}_2 \mid \underline{r})$ collapses to the stationary normal model $l(\underline{\theta}_1 \mid \underline{r})$. It follows that $0 \leq \Lambda \leq 1$. Significance tests for discriminating between these hypotheses can be constructed by noting that the asymptotic distribution of $-2 \log \Lambda$ is chi-square with degrees of freedom equal to the difference in the number of parameters between the two models.

III. Empirical Evidence

A. The Data

The sample consists of daily rate of return data (including all distributions) from July 2, 1962, to December 31, 1980, on each of the 30 stocks in the Dow Jones Industrial Average and the S&P, EW, and VW stock market indexes.[4] The source of this data is the daily stock and index returns files distributed by the Center for Research in Security Prices (CRSP) at the University of Chicago.

B. Departures from the Stationary Normal Hypothesis in the Sample

The statistics presented in Table I assume that the returns for each stock are independent and *identically* distributed.[5] Twenty-six of the 30 stocks have significantly positive skewness statistics (values greater than 0.084). For all stocks in the sample, each kurtosis coefficient has values considerably greater than 3.18, indicating fatter tails than the normal distribution.

[4] Limiting arguments imply that only instantaneous returns, $\ln[(P_t + D_t)/P_{t-1}]$, could be normally distributed. Both the instantaneous return and daily rate of return, $[(P_t + D_t)/P_{t-1}] - 1$, have been used in previous studies. Blattberg and Gonedes [4] use the latter return measure and argue that daily data are consistent with the technical approximation of equivalence between the two. Since their proposed student distribution specification is the most descriptive to date and the benchmark for comparison, the same measure is employed here.

[5] All return statistics are reported in percentages. In the interest of brevity, the serial correlation statistics are not reported. The individual stock returns in the sample conform well to the independence assumption. However, with nonsynchronous trades the observed returns on indexes are serially correlated. The indexes are still included in this study for comparisons with others, evidence on their parameters' stationarity, the stability property of stocks comprising each index, and documentation of the negative skewness in the equally weighted index. Given the positive skewness of the other indexes, this negative skewness must be due to either small firms, the addition of AMEX stocks, or both. Stocks or indexes with complete return data have 4,639 observations. The 5 stocks with less than 4,639 had return observations deleted because of missing prices. Both the return of the day of the missing price and the day after were deleted.

The results in Table I clearly reject the stationary normal distribution hypothesis for each stock and index in the sample. The generalized discrete mixture of normal distributions, however, may be able to explain these results.

C. Model Specification Tests for the Discrete Mixture of Normal Distributions

For each stock and index in the sample, we consider five potential model specifications: $N = 1, 2, 3, 4$, and 5. Given N, the logarithmic likelihood of the normal mixture was maximized by the modified quadratic hill-climbing algorithm described in Goldfeld and Quandt [15]. A comparison of the stationary normal distribution with a mixture of two normals ($N = 2$) can be made with the $-2 \log \Lambda_{12}$ statistic (see Equation (3)). This statistic has an asymptotically chi-square distribution with 3 degrees of freedom. In order to reject the stationary normal distribution null hypothesis in favor of the mixture of two normals at the 0.01 probability level, the statistic must exceed 11.3. All of the 30 stocks and 3 indexes have $-2 \log \Lambda_{12}$ values substantially greater than 11.3. The actual values range from 290.6365 to 889.6253. Hence, the discrete mixture of two normal distributions is considerably more descriptive of the data generating process than the stationary normal model (see Table 5 in Kon [21] for all of the individual statistics in this subsection).

It may be the case, however, that the data were generated by a mixture of more than two normal distributions. For the optimizations of 15 stocks and 3 indexes that were successful in reaching an interior solution, the statistic $-2 \log \Lambda_{23}$ is used to test the null hypothesis that $N = 2$ against the mixture of three ($N = 3$) normal distributions alternative.[6] Fourteen of the 15 stocks and all 3 indexes have substantially greater values than the 11.3 required to reject $N = 2$ in favor of $N = 3$. Of the 10 stocks and 3 indexes that reached an optimum with the mixture of four ($N = 4$) normal distributions model, all were significantly more descriptive than the mixture of two ($N = 2$) normals model. That is, $-2 \log \Lambda_{24}$ (asymptotically chi-square with 6 degrees of freedom) is considerably greater than 16.8 (0.01 probability level) in all cases. For 6 of the 10 $N = 4$ stocks that had respective optimizations available for $N = 3$ and all 3 indexes, their $-2 \log \Lambda_{34}$ statistics indicate that only 3 of the stocks are significant at the 0.05 probability level and none at the 0.01 level.

The mixture of five ($N = 5$) normal distributions model was attempted on the entire sample. Only for the Standard and Poor's 500 Composite Index was a satisfactory interior optimum obtained. However, a value for $-2 \log \Lambda_{45}$ of 3.4670 indicates little additional descriptive ability.

[6] The unsuccessful attempts continually strayed outside the feasible parameter space by selecting negative values for a λ_i until a prespecified iteration limit was exceeded. This behavior is not inconsistent with a true model specification of a smaller N than is currently being attempted. The failure to converge is not, however, a criterion for classifying a stock's return series as being generated by the smaller N specification since there always exists an interior solution of higher N that has the smaller N specification as a subset (e.g., some λ's are zero or there are two or more distributions with the same mean and variance). The inference tests for model specification indicate that if the true model is a mixture of normal distributions, N is likely to be less than five. The inference tests (see Kon [21], Table 5) exhibit significantly declining values for the marginal increments in the log-likelihood available for each stock from $N = 2$ to 3 to 4.

The evidence indicates that with a 0.05 significance level criterion, the sample may be described by a mixture of four normal distributions for 7 stocks, a mixture of three normals for 11 stocks, and a mixture of two normal distributions for the remaining 12 stocks. All 3 indexes are classified by this criterion as being generated by a mixture of three normal distributions.

D. Parameter Estimates for the Discrete Mixture of Normals Models

The parameter estimates and stationarity tests for the mixture of two, three, and four normal distributions models for each stock and index are reported in Tables 6–10 in Kon [21]. For brevity, only summary statistics will be reported here. For all stocks and indexes in the sample there is at least one mean parameter estimate that is negative, and many are statistically significant. Since the day of the week effect reported by French [13] and Gibbons and Hess [14] is a subset of the general mixture of normals model, the negative mean return estimate, for one distribution in 29 out of the 30 stocks and all indexes, is consistent with their Monday effect. However, the estimate of the proportion of observations (λ) associated with that negative mean is considerably larger than the 0.20 predicted by just the Monday effect. Monday returns also exhibited a higher variance than the other days of the week, while the negative mean estimate was generally associated with a lower variance estimate than the variance of the distributions with positive mean estimates. Hence, the true mixture distribution is more complex than a simple partition of the data by day of the week.

Table II summarizes the individual difference tests of the stationarity hypothesis for each stock and index for the mixture of two, three, and four normal distributions models, respectively. For the mixture of two normal distributions model, 27 of the 30 stocks and 2 of the 3 indexes individually rejected the mean stationarity null hypothesis at the 0.05 probability level. For the mixture of three normal distributions model, 12 stocks and all indexes had at least one pair of means (α's) that were significantly different. For the mixture of four normals model, 7 stocks and all indexes had at least one pair of α's that were significantly

Table II

Summary Statistics on Individual Parameter Differences[a]

		Rejections of the Null Hypotheses[b]		
		Mean	Variance	Parameter Vector[c]
Model (N)	No. of Optimum Reached	(At Least One Pair of $\alpha_i - \alpha_j = 0$)	(At Least One Pair of $\sigma_i^2 - \sigma_j^2 = 0$)	(At Least One Pair of $\underline{\gamma}_i - \underline{\gamma}_j = \underline{0}$)
2	30 (3)	27 (2)	30 (3)	30 (3)
3	15 (3)	12 (3)	15 (3)	15 (3)
4	10 (3)	7 (3)	10 (3)	10 (3)

[a] The numbers in parentheses are for the sample of 3 indexes. The numbers without parentheses are for the sample of 30 stocks.

[b] At the 0.05 probability level.

[c] $\underline{\gamma}_i = (\alpha_i, \sigma_i^2)$.

different. Hence, the persistent evidence of significant mean nonstationarity is consistent with the observed skewness in Table I.

The evidence of shifts in the variance parameter in Table II can be used to explain the observed kurtosis in Table I. At least one pair of variance estimates for all stocks and indexes for all models was significantly different. This greater frequency of rejection of stationarity for the variance parameter than the mean indicates its larger contribution to the discrete mixture model specification. An additional test of the stationary model specification is the third null hypothesis in Table II. That is, the nonstationarity of either parameter is sufficient to reject the stationary normal null hypothesis.

The individual parameter stationarity tests and the model specification tests strongly support the discrete mixture of normal distributions as a statistical model of stock returns. The final test of descriptive ability is a comparison with the student distribution model.

E. The Student Distribution

The student model views each return observation as an independent drawing from an identical student density function

$$g(r_t \mid \underset{\sim}{\delta}) = \frac{\Gamma\left(\frac{1+\upsilon}{2}\right) \upsilon^{\upsilon/2}\sqrt{H}}{\Gamma(1/2)\Gamma\left(\frac{\upsilon}{2}\right)} [\upsilon + H(r_t - m)^2]^{-((\upsilon+1)/2)} \tag{4}$$

where m is the location parameter, H is the scale parameter, and υ is the degrees of freedom parameter. The parameter vector, $\underset{\sim}{\delta} = (m, H, \upsilon)$ is to be estimated subject to the parameter space $\psi = \{\underset{\sim}{\delta}: -\infty < m < \infty, 0 < H < \infty, \upsilon > 0\}$. For a sample of size T, the likelihood function for the student distribution is:

$$l(\underset{\sim}{\delta} \mid \underset{\sim}{r}) = \prod_{t=1}^{T} g(r_t \mid \underset{\sim}{\delta}). \tag{5}$$

The maximum likelihood estimator $\hat{\underset{\sim}{\delta}}_T$ of $\underset{\sim}{\delta}$ can be obtained by the supremum of the logarithmic likelihood,

$$L_T(\underset{\sim}{\delta}) = \sum_{t=1}^{T} \log g(r_t \mid \underset{\sim}{\delta}). \tag{6}$$

However, the elements of the vector of first partials of $L_T(\underset{\sim}{\delta})$ with respect to $\underset{\sim}{\delta}$ are nonlinear in $\underset{\sim}{\delta}$ so that gradient or search methods are required to solve for $\underset{\sim}{\delta}$.[7] The location, scale, and degrees of freedom parameter estimates are reported in Table III for each stock and index. The degrees of freedom (υ) estimates range from 3.1177 to 5.5415 which are consistent with an explanation of the observed kurtosis reported in Table I. The student distribution approaches the normal as υ gets large. Therefore, in order to explain the observed magnitude of the fat tails relative to the normal, the degrees of freedom parameter for the student model

[7] Both the modified quadratic hill-climbing algorithm (gradient method) and the pattern direct search method described in Goldfeld and Quandt [15] were used to provide the best possible results for the student model.

Table III

Maximum Likelihood Parameter Estimates for the Student Distribution Model

I.D.	Security (or index)	Location	Scale	Degrees of Freedom
1	Allied Chemical Corp	−0.0050	0.6598	3.9548
2	Aluminum Co America	0.0067	0.6272	4.7871
3	American Brands Inc	0.0100	1.1766	3.8765
4	American Can Co	0.0055	1.4595	3.7784
5	American Tel & Teleg	−0.0141	2.4795	3.6633
6	Bethelem Steel Corp	−0.0282	0.6768	4.2046
7	Du Pont	−0.0178	1.0809	4.1718
8	Eastman Kodak Co	0.0038	0.7666	4.5205
9	Exxon Corp	0.0376	1.2672	4.9064
10	General Electric Co	0.0215	0.9281	4.4235
11	General Foods Corp	−0.0146	1.0290	3.8647
12	General Motors Corp	0.0085	1.0686	4.4142
13	Goodyear	−0.0023	0.6427	5.5415
14	Inco Ltd	−0.0035	0.7670	3.9757
15	Inter Business Mach	0.0100	0.9268	5.0237
16	Inter Harvester Co	0.0170	0.7514	4.4546
17	Inter Paper Co	0.0068	0.6181	5.2064
18	Johns Manville Corp	0.0007	0.7060	3.6542
19	Merck & Co Inc	0.0258	0.7926	5.1347
20	Minnesota Mng & Mfg	0.0059	0.8763	4.7238
21	Owens Illinois Inc	0.0108	0.7809	4.0869
22	Procter & Gamble Co	0.0050	1.4360	3.9691
23	Sears Roebuck & Co	−0.0137	1.1325	3.8898
24	Standard Oil Co Cal	0.0255	1.0364	3.5464
25	Texaco Inc	0.0149	0.9416	3.9757
26	Union Carbide Corp	−0.0073	0.9426	4.1650
27	United Aircraft Prod	−0.0944	0.2755	3.1941
28	United Sts Stl Corp	−0.0488	0.7477	4.2611
29	Westinghouse Elec Co	−0.0150	0.5990	3.5367
30	Woolworth F W Co	−0.0331	0.6992	3.8476
31	Standard & Poor's 500	0.0259	3.0686	4.1754
32	Value Weighted Index	0.0502	3.3293	3.8672
33	Equal Weighted Index	0.1135	3.8711	3.1177

should be in the range $2 < v < 10$ (see Blattberg and Gonedes [4]). The scale (H) estimate is inversely related to the variance of the distribution, $\sigma^2 = \frac{1}{v-2} \cdot \frac{v}{H}$ for $v > 2$. Therefore, the variances of the indexes are, as expected, considerably less than the stocks. The student model is symmetric about its location parameter and hence is unable to explain the significant skewness reported in Table I.

F. Comparison of the Student and Discrete Mixture of Normal Distributions Models

For comparisons between the student model and any specific mixture of N normals model a simple likelihood-ratio is

$$\Lambda_{NS} = \frac{l(\underline{\theta}_N \mid \underline{r})}{l(\underline{\delta} \mid \underline{r})}. \tag{7}$$

The denominator is the student likelihood from Equation (5) and the numerator is the mixture of N normals likelihood from Equation (2). Since the two hypotheses in the likelihood ratio of Equation (7) are not nested in the parameter space as in that of Equation (3), we cannot appeal to the same chi-square distributional result and test the hypotheses. However, we can compare hypotheses by indicating which is more likely. Given the large samples to be used in the following empirical tests, the likelihood ratio in Equation (7) represents the asymptotic posterior odds of the mixture of N normals model to the student (assuming equal prior probabilities).[8] If the log-likelihood ratio (log Λ_{NS}) is positive, the mixture of N normal distributions model is "more likely" than the student model to have generated the stock return data. Alternatively, if the log Λ_{NS} is negative, the student model is more descriptive of the data generating process. A simulation was employed to verify these inferences. Thirty time series of 4,640 observations each were generated assuming a mixture of two normal distributions. Another 30 time series were generated assuming a student distribution. The parameter values selected for the simulation are listed in Table IV, Part A. These values are consistent with those found on actual common stock data. The results in Table IV, Part B are consistent with the interpretation of the log-likelihood ratio above.

Table V contains the computed values of the log Λ_{NS} for $N = 1, 2, 3, 4$, and 5 from the *actual* returns on the 30 stocks and 3 indexes. The log Λ_{1S} provides a comparison of the stationary normal distribution model and the student model. The values of log Λ_{1S} range from -139.777 to -448.294. Hence, the log-odds are substantially in favor of the student model. This result was expected since the student model can account for the kurtosis (fat tails) relative to the stationary normal; but this observed kurtosis is also explained by the shifting variance parameter in the mixture of two or more normal distributions models in Table II. This latter model, however, also has the advantage of being able to explain the observed skewness in Table I. Therefore, the interesting comparison is contained in values of log Λ_{NS} for $N = 2, 3, 4$ and 5.

For all available comparisons with a mixture of three or more normal distributions, the evidence strongly supports the discrete mixture model over the student model. For the 19 stocks and 3 indexes with available values of log Λ_{NS}, $N = 3, 4$, or 5, all are positive. The smallest log-odds value of 6.993 represents odds of 1,089 to 1 in favor of the mixture of normals model. The largest log-odds values of 38.297 corresponds to odds of 4.29×10^{16} to 1. Of the remaining 11 stocks with only values of log Λ_{2S} available for comparison, the evidence is mixed. Four of the 11 stocks strongly support the mixture of two normal distributions model with positive log Λ_{2S} values ranging from 3.557 to 27.471. The remaining 7 stocks support the student model with negative log Λ_{2S} values ranging from -0.661 to -24.261.

In summary, 23 out of the 30 stocks and all 3 indexes have return series that can be better described by a discrete mixture of normal distributions than by the student model. The remaining 7 stocks have series that favor the student model. Note that for these 7 stocks there was no available interior solution to the

[8] See Zellner [30, pp. 291–98].

Table IV

A. True Parameters for the Distribution of Simulated Data

Model	Parameters		
Mixture of Two Normal Distributions	$\alpha_1 = -0.1$	$\sigma_1^2 = 1.0$	$\lambda_1 = 0.65$
	$\alpha_2 = 0.4$	$\sigma_2^2 = 4.0$	$\lambda_2 = 0.35$
Student Distribution	$m = 0.0$	$H = 1.0$	$v = 4.0$

B. Log-Likelihood Ratios for Simulated Data

Replication Number	Student Data	Two Normals Data
1	−21.699	27.613
2	−31.734	12.809
3	−21.090	23.207
4	−13.508	31.664
5	−27.691	36.891
6	−49.172	24.410
7	−27.453	34.422
8	−33.637	28.145
9	−21.215	30.297
10	−11.289	31.605
11	−3.195	20.805
12	−35.617	36.602
13	−1.059	35.316
14	−32.457	19.684
15	−57.570	34.094
16	−35.023	23.699
17	−24.188	35.855
18	−18.102	26.820
19	−34.176	32.133
20	−10.645	27.055
21	−57.395	21.457
22	−26.883	29.500
23	−12.406	48.832
24	−31.680	37.434
25	−13.332	18.988
26	−5.664	29.574
27	−6.832	35.387
28	−31.180	25.367
29	−10.875	31.953
30	−21.855	25.320

mixture of three or more normal distributions models. The prohibitive computer costs for this large a sample allowed us only one attempt per mixture specification for each stock. The strong evidence in favor of the mixture of normals models on the other stocks, where solutions were obtained, indicates that we should be cautious about concluding that these 7 stocks are better described by the student model.

IV. Summary of the Results and Implications

This paper proposed and tested a general formulation and maximum likelihood estimation procedure for a discrete mixture of normal distributions model of

Table V

Log-Likelihood Ratios for Comparing the Student and Discrete Mixture of Normal Distributions Models[a]

I.D.	$\log \Lambda_{1S}$	$\log \Lambda_{2S}$	$\log \Lambda_{3S}$	$\log \Lambda_{4S}$	$\log \Lambda_{5S}$
1	−280.981	−2.959	N.A.[b]	N.A.	N.A.
2	−198.195	−13.396	12.717	19.665	N.A.
3	−300.484	−6.342	11.370	13.759	N.A.
4	−266.721	−0.669	10.883	N.A.	N.A.
5	−319.014	−6.062	27.825	N.A.	N.A.
6	−277.838	−16.080	N.A.	N.A.	N.A.
7	−213.120	10.901	23.679	28.072	N.A.
8	−210.790	−6.935	14.799	20.250	N.A.
9	−145.093	10.085	13.776	N.A.	N.A.
10	−211.594	−1.382	N.A.	9.933	N.A.
11	−254.476	−0.777	19.233	N.A.	N.A.
12	−214.106	−2.698	N.A.	9.459	N.A.
13	−139.777	5.541	N.A.	N.A.	N.A.
14	−239.213	3.557	N.A.	N.A.	N.A.
15	−169.473	−4.276	13.978	N.A.	N.A.
16	−195.641	−6.447	13.677	N.A.	N.A.
17	−161.734	−0.661	N.A.	N.A.	N.A.
18	−283.539	−3.214	N.A.	23.690	N.A.
19	−163.584	−3.445	14.108	N.A.	N.A.
20	−195.043	−1.681	N.A.	N.A.	N.A.
21	−257.204	−10.376	6.993	N.A.	N.A.
22	−272.488	−10.140	10.765	13.250	N.A.
23	−282.283	−4.338	N.A.	N.A.	N.A.
24	−282.078	12.647	N.A.	N.A.	N.A.
25	−251.143	1.166	N.A.	12.502	N.A.
26	−226.607	−2.331	17.827	N.A.	N.A.
27	−385.098	27.471	N.A.	N.A.	N.A.
28	−311.072	−16.484	31.712	34.371	N.A.
29	−444.849	−24.261	N.A.	N.A.	N.A.
30	−306.896	−6.566	N.A.	N.A.	N.A.
31	−222.506	−4.701	20.457	21.485	23.219
32	−250.715	0.543	27.575	29.601	N.A.
33	−448.294	−3.482	35.848	38.297	N.A.

[a] The log-likelihood ratio, $\log \Lambda_{NS}$, is the logarithm of the ratio of the likelihood for the discrete mixture of N normal distributions divided by the likelihood for the student distribution. Negative values support the student distribution hypothesis while positive values support the mixture of N normal distributions hypothesis.

[b] N.A. stands for not available.

stock returns. The likelihood ratio tests of model specification indicate that the sample of 30 stocks can be described by a mixture of four normal distributions for 7 stocks, a mixture of three normals for 11 stocks, and a mixture of two normal distributions for the remaining 12 stocks. All 3 indexes in the sample can be described by a mixture of three normal distributions. Stationarity tests on the parameter estimates of the discrete mixture of normal distributions model revealed significant differences in the mean estimates that can explain the observed skewness in security returns. Significant differences in the variance estimates can explain the observed kurtosis. Furthermore, the discrete mixture of normal distributions model has substantially more descriptive validity than the student model.

These results have several implications for theoretical models and empirical research in financial economics. For the single-period mean-variance consumer equilibrium model and two-parameter capital asset pricing models, the evidence is consistent with the normality assumption. It is for the multiperiod theoretical models (i.e., option pricing) and empirical research that the nonstationarity evidence implies severe violations of model specification and estimation procedures.

The application of the discrete mixture of normal distributions model to empirical research requires that each rate of return observation be classified according to its respective normal distribution. Then the appropriate parameter estimates and sampling distribution can be used to construct inference tests. Given the parameter estimates of the discrete mixture from the maximum likelihood procedure, the "most likely" classification rule is

$$\operatorname*{Max}_{i} \hat{\lambda}_i p(r_t \mid \hat{\underline{\gamma}}_i). \tag{8}$$

That is, select the distribution i for generating observation t that has the largest posterior probability.[9] Note that in this procedure the parameter shift dates have also been estimated without a priori knowledge of the event dates. This may be particularly useful for efficient markets tests when the estimated data partition can be associated with corresponding public announcements or information signals in accounting numbers released prior to the event.[10] Furthermore, if the information also results in a change in equilibrium expected returns, the estimates of the true mean and variance for *each* day surrounding the event is available to construct appropriate inference tests of abnormal performance.

The classification procedure was used on the stock and index returns sample in this paper to test whether the classified subsets of data conformed to the normality assumption. *None* of the kurtosis tests for any of the subsets of each stock and index exhibited the fat tail property. Hence, this paper's view (and Christie [9]) that the mixture of normals is being generated by parameter nonstationarity is empirically verified.

The methodology described above can be particularly useful in identifying the mean and variance changing process for stocks and indexes. Knowledge of whether a security or index's nonstationary process is dominated by either cyclical or time-ordered shifts is crucial to the problem of predicting parameters. Furthermore, if cyclical shifts, like the day of the week effect, dominate the sample this will be consistent with the evidence that monthly stock returns are approximately normally distributed.

Another area for further research is the identification of macro versus micro (firm-specific) information. Suppose the total return variance of a stock is defined

[9] This procedure was proposed by Kon and Lau [23] for mixtures of linear models and has been applied by Kon [20] and Christie [9].

[10] See Brown and Warner [6] for evidence on the power of tests when the announcement date is unknown.

by the market model as $\sigma_t^2(r) = \beta_t^2\sigma_t^2(R_M) + \sigma_t^2(\epsilon)$. Then the source of total return variance nonstationarity may be due to changes in the stock's beta and/or residual variance (micro effects) or changes in the market's return variance (macro effect) or both. Clearly, the significant nonstationarity of the market's return variance in this paper implies that it is an important component for high beta stocks. Finally, empirical evidence on the variance changing process for the market portfolio can be useful in the application of Merton's [24] estimator for the expected return on the market portfolio.

Appendix

Parameter Estimation for the Discrete Mixture of Normal Distributions

Given the T observations on the stock return variable, r_t, there exists some *true* permutation of the rows of $\underset{\sim}{r} = (r_1, r_2, \cdots, r_T)'$ which will allow them to be partitioned according to Equation (1). Assume a multinomial prior whereby the equation associated with set I_i is selected for generating observations with probability λ_i, $i = 1, 2, \cdots, N$. Therefore, each observation is viewed as a drawing from a mixture distribution (conditional probability density function [p.d.f.] of r_t given the parameter vector $\underset{\sim}{\theta}$),

$$f_t(r_t \mid \underset{\sim}{\theta}) = \sum_{i=1}^{N} \lambda_i p(r_t \mid \underset{\sim}{\gamma}_i) \tag{A1}$$

where $\underset{\sim}{\gamma}_i = (\alpha_i, \sigma_i^2)$, $i = 1, 2, \cdots, N$: $\underset{\sim}{\theta} = (\underset{\sim}{\gamma}_1, \underset{\sim}{\gamma}_2, \cdots, \underset{\sim}{\gamma}_N, \lambda_1, \lambda_2, \cdots, \lambda_{N-1})$ is the parameter vector to be estimated subject to the parameter space $\Omega = \{\underset{\sim}{\theta}: -\infty < \alpha_i < \infty, 0 < \sigma_i^2 < \infty, i = 1, 2, \cdots, N, 0 < \lambda_i < 1, i = 1, 2, \cdots, N-1\}$; $p(r_t \mid \underset{\sim}{\gamma}_i)$ is a normal probability density function with mean α_i and variance σ_i^2, $i = 1, 2, \cdots, N$; $\sum_{i=1}^{N} \lambda_i = 1$; $\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_N^2$; and $T_i \geq 2$, $i = 1, 2, \cdots, N$.[11]

Estimation of the parameter vector $\underset{\sim}{\theta}$ can be done by choosing values that maximize the likelihood function

$$l(\underset{\sim}{\theta} \mid \underset{\sim}{r}) = \prod_{t=1}^{T} \left[\sum_{i=1}^{N} \lambda_i p(r_t \mid \underset{\sim}{\gamma}_i)\right]. \tag{A2}$$

The generality of this likelihood function can be seen by comparing it to the likelihood function assuming that the parameters and partition of the T observations into T_i, $i = 1, 2, \cdots, N$ were *known a priori*. This true likelihood function would be

$$l(\underset{\sim}{\theta} \mid \underset{\sim}{r}) = \prod_{t=1}^{T_1} p(r_t \mid \underset{\sim}{\gamma}_1^*) \prod_{t=1}^{T_2} p(r_t \mid \underset{\sim}{\gamma}_2^*) \cdots \prod_{t=1}^{T_N} p(r_t \mid \underset{\sim}{\gamma}_N^*) \tag{A3}$$

where no proportionalities are necessary. However, since knowledge of this partition is *un*available a priori, we cannot employ the true likelihood function directly. However, we can show that maximizing the proposed likelihood will achieve the same result. The right-hand side of the proposed likelihood function

[11] The requirement that there be at least two observations in each set is necessary to define a positive variance. The strict ordering of variances insures that the mixture is uniquely determined. The information event scenario previously discussed can be used to justify this condition. See Kon and Jen [22] for a discussion and simulation of the importance of this retriction.

(Equation (A2)) can be expanded to obtain

$$l(\underline{\theta} \mid \underline{r})$$

$$= \sum_{\substack{T_1=0 \\ (T_1+T_2+\cdots+T_N=T)}}^{T} \sum_{T_2=0}^{T} \cdots \sum_{T_N=0}^{T} \frac{T!}{T_1! \cdots T_N!} \lambda_1^{T_1} \lambda_2^{T_2} \cdots \lambda_N^{T_N} h_{(T_1,\cdots,T_N)}(\underline{\gamma}_1, \cdots, \underline{\gamma}_N) \quad \text{(A4)}$$

where

$$h_{(T_1,T_2,\cdots,T_N)}(\underline{\gamma}_1, \underline{\gamma}_2, \cdots, \underline{\gamma}_N)$$

$$= \prod_{t=0}^{T_1} p(r_t \mid \underline{\gamma}_1) \prod_{t=0}^{T_2} p(r_t \mid \underline{\gamma}_2) \cdots \prod_{t=0}^{T_N} p(r_t \mid \underline{\gamma}_N). \quad \text{(A5)}$$

The summation in the likelihood Equation (A4) is over all feasible values of T_1, T_2, $\cdots$ and T_N and, given each set of values for T_1, T_2, $\cdots$, T_N, over all possible ways of partitioning the $\underline{r}$ vector into the vectors $\underline{r}_{(T_i)}$, $i = 1, 2, \cdots, N$ with T_i, $i = 1, \cdots, N$ observations in each vector, respectively. For example, let $a_{(T_1,T_2,\cdots,T_N)}$ be the event that a particular T_1 of the T r_t's is from $p(r_t \mid \underline{\gamma}_1)$, T_2 of the T r_t's is from $p(r_t \mid \underline{\gamma}_2)$, $\cdots$, and T_N of the T r_t's is from $p(r_t \mid \underline{\gamma}_N)$. The prior probability of the event $a_{(T_1,T_2,\cdots,T_N)}$ is $p^{(T_1,T_2,\cdots,T_N)} = \lambda_1^{T_1}\lambda_2^{T_2} \cdots \lambda_N^{T_N}$ corresponding to only one way of partitioning the r vector into $\underline{r} = (\underline{r}_{(T_1)}, \underline{r}_{(T_2)}, \cdots, \underline{r}_{(T_N)})'$. The same prior probability, $\lambda_1^{T_1}\lambda_2^{T_2} \cdots \lambda_N^{T_N}$, is associated with $T!/(T_1!\ T_2! \cdots T_N!)$ possible partitions. However, the sample information (likelihood factor), $h_{(T_1,T_2,\cdots T_N)}(\underline{\gamma}_1, \underline{\gamma}_2, \cdots, \underline{\gamma}_N)$, associated with each of these paritItions is different. Note that given the specific paritition T_1, T_2, $\cdots$, T_N, the values of the parameters $\underline{\gamma}_1$, $\underline{\gamma}_2$, $\cdots$, $\underline{\gamma}_N$ that maximize $h_{(T_1,T_2,\cdots,T_N)}(\underline{\gamma}_1, \underline{\gamma}_2, \cdots, \underline{\gamma}_N)$ will be identical to the parameter values $\underline{\gamma}_1^*$, $\underline{\gamma}_2^*$, $\cdots$, $\underline{\gamma}_N^*$ in Equation (A3) that maximize the true likelihood function. Furthermore, since the N^T individual terms of the summation in the proposed likelihood function (Equation (A4)) represent the complete posterior distribution, a large sample property is that the likelihood factor, $h_{(T_1T_2,\cdots,T_N)}(\underline{\gamma}_1, \underline{\gamma}_2, \cdots, \underline{\gamma}_N)$, will dominate the posterior p.d.f.[12] Then as all values of T_1, T_2, $\cdots$, T_N are spanned, all N^T possible combinations are evaluated. Hence, the parameter estimates of $\underline{\gamma}_i$, $i = 1, 2, \cdots, N$ and $\lambda_i = \frac{T_i}{T}$, $i = 1, 2, \cdots, N-1$ that maximize the likelihood function in Equations (A2) or (A4) are the same for those that maximize the *true* likelihood (Equation (A3)). Therefore, any scenario that generates data from N normal probability distributions *in any order* is included as a subset of this general specification.

For a given N and T, the maximum likelihood estimator $\hat{\underline{\theta}}_T$ of $\underline{\theta}$ is defined implicitly by the supremum of the logarithmic likelihood,

$$L_T(\underline{\theta}) = \sum_{t=1}^{T} \log f_t(r_t \mid \underline{\theta}). \quad \text{(A6)}$$

Then the maximum likelihood estimator $\hat{\underline{\theta}}_T$ is the solution to the likelihood equations

$$\frac{\partial L_T(\underline{\theta})}{\partial \underline{\theta}} = \sum_{t=1}^{T} \frac{1}{f_t(r_t \mid \underline{\theta})} \cdot \frac{\partial f_t(r_t \mid \underline{\theta})}{\partial \underline{\theta}} = \underline{0} \quad \text{(A7)}$$

[12] See Zellner [30, pp. 31–3] for a proof.

such that the matrix of second partials evaluated at the solution point is negative definite. Note that the elements of $\partial L_T(\underline{\theta})/\partial \underline{\theta}$ in Equation (A7) are nonlinear in $\underline{\theta}$ so that no explicit closed form solution for $\hat{\underline{\theta}}_T$ exists. However, Equation (A7) represents an implicit solution(s) for $\hat{\underline{\theta}}_T$. Therefore, in practice, we can employ iterative gradient methods that find all solutions (local maxima) and choose the maximum maximorum.[13]

Hypothesis tests on the elements of the estimated parameter vector, $\hat{\underline{\theta}}_T$, can be constructed with the information in the sample covariance matrix. For maximum likelihood estimates, the sample covariance matrix is the negative inverse of the matrix of the second partial derivatives of the logarithmic likelihood function with respect to the parameter vector evaluated at $\underline{\theta} = \hat{\underline{\theta}}_T$.

[13] Kiefer [19] has verified that the likelihood equations do have a consistent root in Ω.

REFERENCES

1. Joseph Aharony and Itzhak Swary. "Quarterly Dividend and Earnings Announcements and Stockholders' Returns: An Empirical Analysis." *Journal of Finance* 35 (March 1980), 1–12.
2. Clifford Ball and Walter Torous. "A Simplified Jump Process for Common Stock Returns." *Journal of Financial and Quantitative Analysis* 18 (March 1983), 53–65.
3. William Beaver. "The Information Content of Annual Earnings Announcements." *Empirical Research in Accounting: Selected Studies, Supplement to Journal of Accounting Research* 6 (1968), 67–92.
4. Robert Blattberg and Nicholas Gonedes. "A Comparison of the Stable and Student Distributions as Statistical Models for Stock Prices." *Journal of Business* 47 (April 1974), 244–80.
5. A. James Boness, Andrew Chen, and Som Jatusipitak. "Investigations of Nonstationarity in Prices." *Journal of Business* 47 (October 1974), 518–37.
6. Stephen Brown and Jerold Warner. "Measuring Security Price Performance." *Journal of Financial Economics* 8 (September 1980), 205–58.
7. Guy Charest. "Dividend Information, Stock Returns, and Market Efficiency—II." *Journal of Financial Economics* 6 (June/September 1978), 297–330.
8. Peter Clark. "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices." *Econometrica* 41 (January 1973), 135–55.
9. Andrew Christie. "On Information Arrival and Hypothesis Testing in Event Studies." Working paper, University of Rochester, 1983.
10. ———. "The Stochastic Behavior of Common Stock Variances: Value, Leverage, and Interest Rate Effects." *Journal of Financial Economics* 10 (December 1982), 407–32.
11. H. David, H. Hartley, and E. Pearson. "The Distribution of the Ratio, in a Single Normal Sample, of Range to Standard Deviation." *Biometrika* 61 (1954), 491.
12. Eugene Fama. "The Behavior of Stock Prices." *Journal of Business* 47 (January 1965), 244–80.
13. Kenneth French. "Stock Returns and the Weekend Effect." *Journal of Financial Economics* 8 (March 1980), 55–70.
14. Michael Gibbons and Patrick Hess. "Day of the Week Effects and Asset Returns." *Journal of Business* 54 (October 1981), 579–96.
15. Stephen Goldfeld and Richard Quandt. *Nonlinear Method in Econometrics.* Amsterdam: North-Holland, 1972.
16. Robert Hagerman. "More Evidence on the Distribution of Security Returns." *Journal of Finance* 33 (September 1978), 1213–20.
17. Der-Ann Hsu, Robert Miller, and Dean Wichern. "On the Stable Paretian Behavior of Stock-Market Prices." *Journal of the American Statistical Association* 69 (March 1974), 108–13.
18. Donald Keim. "Size Related Anomalies and Stock Return Seasonality: Further Empirical Evidence." Working paper, University of Chicago, Graduate School of Business, 1981.
19. Nicholas Kiefer. "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model." *Econometrica* 46 (March 1978), 427–34.

20. Stanley J. Kon. "The Market Timing Performance of Mutual Fund Managers." *Journal of Business* 56 (July 1983), 323–47.
21. ———. "Models of Stock Returns—A Comparison." Salomon Brothers Center for the Study of Financial Institutions. Working paper, Series No. 258, New York University, 1982.
22. ——— and Frank C. Jen. "The Investment Performance of Mutual Funds: An Empirical Investigation of Timing, Selectivity, and Market Efficiency." *Journal of Business* 52 (April 1979), 263–89.
23. ——— and W. Patrick Lau. "Specification Tests for Portfolio Regression Parameter Stationarity and the Implications for Empirical Research." *Journal of Finance* 34 (May 1979), 451–67.
24. Robert Merton. "On Estimating the Expected Return on the Market: An Exploratory Investigation." *Journal of Financial Economics* 8 (December 1980), 323–61.
25. Robert Officer. "The Distribution of Stock Returns." *Journal of the American Statistical Association* 67 (December 1972), 807–12.
26. James Patell and Mark Wolfson. "The Ex Ante and Ex Post Price Effects of Quarterly Earnings Announcements Reflected in Option and Stock Prices." *Journal of Accounting Research* 19 (Autumn 1981), 434–58.
27. Peter Praetz. "The Distributions of Share Price Changes." *Journal of Business* 45 (January 1972), 49–55.
28. S. James Press. "A Compound Events Model of Security Prices." *Journal of Business* 40 (July 1967), 317–35.
29. Richard Roll and Stephen Ross. "An Empirical Investigation of the Arbitrage Pricing Theory." *Journal of Finance* 35 (December 1980), 1073–1103.
30. Arnold Zellner. *An Introduction to Bayesian Inference in Econometrics.* New York: John Wiley and Sons, 1971.