Sequential Parameter Nonstationarity in Stock Market Returns

DONGCHEOL KIM
Department of Finance, School of Business, Rutgers University, New Brunswick, NJ 08903

STANLEY J. KON
Department of Finance, School of Business Administration, University of Michigan, Ann Arbor, MI 48109

Abstract. This paper provides a Bayesian test of parameter nonstationarity and an estimation procedure for the detection of change points in the time series of stock returns. The empirical results indicate that this procedure can identify the change points in the data without prior knowledge and provide substantially more descriptive validity for the distribution of stock returns than competing models.

Key words: parameter nonstationarity, change points, stock return distribution, model comparisons

1. Introduction

The form of the distribution for security returns plays an important role in financial market theory and practice. A statistical distribution assumption concerning the financial securities available for wealth allocation is required for portfolio selection models and in equilibrium asset pricing. Furthermore, multiperiod models of asset pricing and contingent claims also require knowledge of the time path for the parameters of the statistical distribution assumed.

The most convenient assumption for both theoretical and empirical research is that the return series of an individual security are independent drawings from an identical (stationary) normal distribution. Both theoretical and empirical evidence, however, strongly suggest that a time series of stock returns will contain significant parameter nonstationarity. In a dynamic economic environment, the investment and financing decisions of firms will affect the systematic risk, expected return, and standard deviation of equity returns. For example, Boness et al. (1974) find parameters shift after a capital structure change and Christie (1982) demonstrates that the standard deviation of a stock's return is an increasing function of both financial and operating leverage. Beaver (1968), with realized returns, and Patell and Wolfson (1981), with ex ante assessments, find an increase in the variance of stock returns around the announcement of quarterly earnings.

Macroinformation shocks may also shift the level of interest rates and market risk premiums. For example, Officer (1972) and Hsu et al. (1974) find a substantial increase in the characteristic exponent of common stocks from the pre-World War II period to the postwar period. Therefore, it is not surprising for the expected return and the risk dimensions of individual stocks and portfolios to change through time. In fact, the often observed skewness, more highly peaked and longer-tailed (leptokurtic) characteristics of the empirical

unconditional distribution of daily stock returns are consistent with the nonstationarity of the mean and standard deviation parameters. The difficulty, however, is modeling and estimating the changing parameter values. Therefore, researchers have proposed a variety of alternative explanations in the form of a mixture of normal distributions.

Blattberg and Gonedes (1974) consider two continuous mixtures of normal distribution models. By assuming that the variance of the normal distribution follows an inverted gamma distribution, the resulting posterior distribution is the Student. They also show that the nonnormal stable Paretian distribution hypothesis proposed by Mandelbrot (1963a, 1963b, 1967) and Fama (1965) is obtained when the variance of the normal distribution follows a strictly positive stable distribution with characteristic exponent between zero and one. The evidence in Blattberg and Gonedes indicates that the Student distribution model provides a better empirical description of stock returns than the stable Paretian model.[1]

Christie (1983) has formulated a discrete mixture of two normal distribution models where returns drawn from the higher variance represent information events while the other distribution generates noninformation events. Ball and Torous (1988) use a variant of this model to approach the problem of single event date uncertainty. A generalized discrete mixture of normal distributions has been proposed and tested for daily stock returns in Kon (1984). Each return observation is viewed as a drawing from one of a finite number of normal distributions, with some fixed probability. This model refers to subsets of the data whose observations are not necessarily consecutive in time so that both structural and cyclical parameter shifts can be accommodated. The January effect or the weekend effect might be examples of cyclical parameter shifts. The release of relevant firm-specific information (e.g., a merger, capital structure change, etc.) might result in a structural parameter shift. This model formulation and estimation procedure provides both generality and a better statistical description of stock returns than the Student model. Finally, Press (1968), Jarrow and Rosenfeld (1984), Ball and Torous (1985), and Akgiray and Booth (1986, 1987) have provided some positive evidence on a Poisson jump-diffusion process for the distribution of stock returns. This model is essentially an infiniite mixture of normal distributions with parameter restrictions.

This paper focuses on the empirical issue of which mixture of independent normal distributions model best describes actual daily stock return data. The resolution of this fundamental empirical issue is necessary for building relevant theoretical models and for the application of well-specified empirical tests of asset pricing models and the efficient markets hypothesis. The maximum likelihood estimation procedure in all of the previous models provide estimates of all the parameters without identifying each observation with its respective generating distribution. The purpose of this paper is to demonstrate the benefits of identification of observations with their respective generating distribution. A more descriptive model of the distribution of stock returns is obtained by adding minimal structure to the model specification and employing a Bayesian estimation procedure. The additional structure is a sequential (or time-ordered) mixture of normal distributions model. That is, a sequence of return observations are normal variates with stationary parameters up to a change point. The sequential structure requires that at least two consecutive observations are generated from the same distribution to define a regime with positive variance before another shift can take place. With this minimal requirement, a Bayesian detection method for the unknown change points $(\tau_1, \tau_2, \ldots)$ is derived and tested.

The sequential mixture methodology in this paper can be used to generate and accommodate a large number of information distributions with varying parameter magnitudes and multiple event date uncertainty. This is important for capturing the effects of multiple announcements with information leakage as may arise. These problems arise in a bidding contest in mergers and acquisitions or when measuring the effect of tax or regulatory policy changes.

The remainder of this paper is organized as follows: In section 2, the model specification for the sequential mixture alternative is defined. The model formulation and estimation of change points methodology are provided. Section 3 contains a description of the competing models. The empirical evidence comparing the alternative models for the distribution of stock returns is presented in section 4. A summary of the results and implications are given in section 5.

2. Sequential mixture of normal distributions model specification

The previous mixture of independent normal distributions models in the literature use maximum likelihood estimation procedures that provide estimates of all the parameters without identifying each observation with its respective generating distribution. This section presents a new sequential mixture model. By adding minimal structure to the model specification and using a Bayesian estimation procedure, this technique can provide the benefits of (1) Identification of observations with their respective generating distribution; (2) computational simplicity; and (3) a more descriptive model of the distribution of stock returns.

2.1. The stationarity normal distribution null hypothesis

One motivation for using a mixture of normal distributions model to describe daily stock return data is skewness and kurtosis relative to a stationary normal assumption. This phenomenon is exhibited in table 1 for our sample of daily rates of return from July 2, 1962 to December 3, 1990 on the 30 stocks in the Dow Jones Industrial Average (DJIA). We also report these statistics on three stock indexes: Standard & Poors 500 (S&P), Center for Research in Security Prices (CRSP), Equally Weighted (EW), and CRSP Value-Weighted (VW). All data comes from the CRSP at the University of Chicago (CRSP tapes).

Note that all individual sticks and indexes have statistics that reject the stationary (identical) normal distribution null hypothesis. The evidence indicates significantly fatter tails than the stationary normal distribution for all individual stocks and indexes. The excess kurtosis statistic ranges in value from 3.3213 to 73.7098 when the upper 1 percentile in the distribution is 0.13. All but five individual stocks reject the symmetric distribution null hypothesis with skewness coefficients greater (or less) than the 1 percentile critical value of 0.058 (-0.058). Most of the distributions are left-tail heavy. Given the relative magnitudes of these statistics, the stationary normality hypothesis is rejected more severely by the excess kurtosis statistics than the skewness statistics.

Table 1. Sample statistics for daily returns (July 2, 1962–December 31, 1990).

I.D. No.	Security (or Index)	Sample Size	Mean (%)	Standard Deviation (%)	Skewness Coefficient	Excess Kurtosis
1	Allied Signal	1,328	0.0117	2.0594	0.2018	73.7098
2	Alcoa	7,167	0.0448	1.7386	−0.3221	7.9657
3	American Exp	3,441	0.0586	2.0386	−0.3778	12.3701
4	AT&T	7,166	0.0415	1.1934	−0.1692	22.6588
5	Bethlehem Stl	7,167	0.0274	2.0640	−0.4456	14.8119
6	Boeing	7,167	0.0854	2.1564	0.5644	4.4869
7	Caterpillar	7,167	0.0456	1.6644	−0.3081	8.2662
8	Chevron	7,167	0.0589	1.5638	0.0450	4.8959
9	Coca Cola	7,167	0.0713	1.5189	−0.1648	17.5454
10	Disney Walt	7,167	0.0972	2.0860	−0.2979	10.2161
11	Du Pont	7,166	0.0403	1.4127	−0.0557	6.4517
12	Eastman Kodak	7,167	0.0458	1.6050	−0.1746	27.5529
13	Exxon	7,167	0.0612	1.2672	−0.3869	24.1442
14	General Electric	7,167	0.0518	1.4474	−0.0290	5.8129
15	General Motors	7,166	0.0411	1.4303	0.0104	10.7273
16	Goodyear	7,167	0.0355	1.7176	−0.2166	14.6313
17	IBM	7,164	0.0462	1.3682	−0.3121	13.8527
18	Int'l Paper	7,167	0.0510	1.6784	−0.2889	12.4147
19	McDonalds	6,158	0.0999	2.0347	0.0418	6.3015
20	Morgan JP	5,495	0.0629	1.5984	−0.3530	53.1677
21	Merck	7,167	0.0739	1.4341	0.1346	3.3408
22	3M	7,167	0.0491	1.4285	−0.5951	18.4713
23	Phillip Morris	7,167	0.0939	1.5605	0.1490	3.3213
24	P&G	7,167	0.0541	1.3079	−0.3051	39.1795
25	Sears	7,167	0.0329	1.5289	−0.0708	16.3683
26	Texaco	7,166	0.0503	1.5328	0.0944	6.8170
27	Union Carbide	7,167	0.0419	1.6307	0.1320	10.0492
28	United Tech.	7,165	0.0591	1.7923	0.1910	3.3245
29	Westinghouse	7,167	0.0644	1.9403	−0.5319	20.9271
30	Woolworth	7,167	0.0584	1.7611	0.6108	7.4724
31	S&P 500	7,167	0.0288	0.8903	−1.6452	44.2041
32	EW CRSP	7,167	0.0700	0.7917	−0.9274	25.1124
33	VW CRSP	7,167	0.0438	0.8435	−1.3757	34.7037

Notes: The upper and lower 1 percentile points in the distribution of the skewness statistic are 0.058 and −0.058, respectively.

The upper and lower 1 percentile points in the distribution of the excess kurtosis statistic are 0.13 and −0.11, respectively.

2.2. *Preliminary evidence on the sequential model*

In an efficient market, public announcements of corporate investment and financial decisions that imply a change in the firm's expected return and risk will be impounded in stock prices immediately. The announcements of relevant macroeconomic information will affect

the return and risk of all securities, and hence, portfolios (indexes). These events may represent change points, and between such events, there may be stationarity.[2] Cyclical (or seasonal) nonstationarity may result from the anticipation of regular quarterly (e.g., earnings and dividend) or weekly (e.g., macroeconomic information) announcements. Sequential nonstationarity generally results from unanticipated announcements like capital structure changes, mergers, stock splits, an oil embargo or market failure (e.g., October 19, 1987). For some intuition on whether the cyclical or time-ordered characteristic of nonstationarity is more dominant, we can partition the data to minimize nonstationarity. Consider the following partitions of the 7,167 daily observations in the stock return series with the number of regimes in parentheses: no partition (1); every five years (6); every three years (10); yearly (29); quarterly (114); by day of the week (5); by month of the year (12); and by year and day of the week (145). For example, the day of the week scenario groups all Monday returns into the first regime, all Tuesday returns into the second regime, and so forth. The last three daily data partitions have cyclical (or seasonal) components. However, with the requirement of at least two consecutive observations to be considered a sequential (time-ordered) model, a partition by month of the year has a major cyclical component that can be estimated sequentially with daily data.

The entries in table 2 represent the number of stocks among the 30 securities and three indexes for which the partition indicated by the column is more plausible than the partition indicated by the row. Each pairwise comparison is performed by means of a χ^2 test at the 5% significance level. The numbers in table 2 indicate that the time-ordered partition by quarter is the most plausible. The yearly partition is the next likely, and the mixture of time-ordered and cyclical components in the yearly and day of the week partition follows. This gives us some intuition that the time-ordered shift of parameters may be more likely than the entirely cyclical partition by day of the week.[3]

Since the time-ordered pattern is more plausible than cyclical, we can expect a lot of overlap between the sequential mixture and generalized discrete mixture specifications. That is, both can accommodate the time-ordered pattern, a priori. Furthermore, we do no know the change-points a priori, but we do expect many such points in a time-ordered pattern. Which model specification and estimation procedure is more powerful in detecting change points and describing the data is an empirical issue.

2.3. Model formulation

The specification of a sequential mixture of normal distributions model is a sequence of stock returns $(R_1, R_2, \ldots, R_T)$ that are normal variates with stationary mean or variance parameter up to a change point. Define these as τ_1 return observations from regime 1. Thereafter, return $\tau_1 + 1$ to τ_2 are a sequence of return observations from regime 2, and so forth, until the last sequence of return observations from $\tau_{K-1} + 1$ to τ_K is the Kth regime. The formulation of this model is as follows:

Regime 1: $R_t = \mu_1 + \epsilon_{1t},$ $t = 1, 2, \ldots, \tau_1$

Regime 2: $R_t = \mu_2 + \epsilon_{2t},$ $t = \tau_1 + 1, \ldots, \tau_2$

$\quad \cdot \qquad \cdot$

Regime K: $R_t = \mu_K + \epsilon_{Kt},$ $t = \tau_{K-1} + 1, \ldots, \tau_K(= T)$

Table 2. Number of securities for which the partitions in the column fit daily returns better than the partitions in the row. The sample consists of 30 stocks and 3 indexes. Pairwise χ^2 tests at the 5% significance level are used.

Methods of Partition	No Partition (1)	Quarterly (114)	Yearly (29)	Every Three Years (10)	Every Five Years (6)	Day of the Week (5)	Month of the Year (12)	Yearly and Day of the Week (145)
No Partition	•	33	33	33	33	33	33	33
Quarterly	0	•	0	0	0	0	1	3
Yearly	0	33	•	0	0	0	2	14
Every three years	0	33	33	•	2	3	10	20
Every five years	0	33	33	31	•	11	33	33
Day of the week	0	33	33	30	22	•	33	33
Month of the year	0	32	31	20	0	0	•	32
Yearly and day of the week	0	30	19	13	0	0	1	•

The numbers in parentheses represent the number of regimes in each regular partition if all data is available over the whole period from July 2, 1962 to December 31, 1990. The number of regimes for the securities having no full sample (I.D. No. 1, 3, 19 and 20) can be determined accordingly.

where, for each k, ϵ_{kt} is (for each $t \in [\tau_{k-1} + 1, \tau_k]$) identically and independently normally distributed with mean 0 and variance σ_k^2. Change points for the mean and change points for the variance are not necessarily coincident. K indicates the total number of regimes in which either a mean or a variance parameter deterministically shifts. Note that parameters $\tau_1, \ldots, \tau_{K-1}$ define the change points. The nonstationarity arises from observations that are drawn from a different population after each change point.

On a posteriori ground, the sequential mixture specification is a special case of the generalized discrete mixture of normal distributions model.[4] However, since knowledge of the partition is unavailable a priori, the a priori formulation of the partition has important implications. Only the sequential mixture specification attains both computational simplicity and observaton identification for a p otentially larger number of regimes.

2.4. A detection procedure for change points

The major task in the sequential mixture of normal distributions model is how to detect change points for the mean and for the variance. There are two ways to detect the change points, simultaneously and sequentially. If the number of regimes, K, is known, the simultaneous detection method selects the set of change points that has the greatest posterior probability (or the likelihood value after adjusting for the number of estimated parameters). This is an optimal method. In reality, however, the number of regimes is not explicitly known. Furthermore, even if the number of regimes are known, there will be situations in which the simultaneous detection of change points is computationally very difficult, in particular when the sample size and the number of regimes are relatively large. This situation is typical in our stock return data. If the sample size is T and the number of regimes is K, approximately $\binom{T}{K-1}$ computations of the joint posterior probability of change points are needed for all possible combinations to be evaluated. For example, if the number of daily return observations is 7,000 and the number of regimes is 101, then approximately 10^{227} computations of the posterior probability are needed. In other words, we have to consider 10^{227} different change points vectors. Among those, we choose one change point vector as the estimate of $\tau_1, \ldots, \tau_{K-1}$ that has the greatest posterior probability. To avoid the difficulties of the simultaneous detection method, we suggest a sequential detection procedure. Although this is theoretically suboptimal, it is computationally more efficient and can estimate the unknown number of change points. The simulation evidence in the next subsection will be used to judge its effectiveness.

Since change points for the mean and for the variance are not necessarily coincident, we estimate change points for the mean and for the variance separately. The sequential change point detection procedure begins with a test of the stationarity of the mean (variance) parameter applied to an initial sample (i.e., the first 30 observations). For the stationarity test of the mean parameter over a given sample period, we consider the following null hypothesis

$$H_{01} : \delta = \mu_1 - \mu_2 = 0,$$

where μ_1 and μ_2 are mean parameters of two implicit regimes in the initial sample. The test requires computation of the unconditional p-value for H_{01} unconditionally on any variance

parameter change. This stationarity test of the mean parameter is repeated as each data point is added to the initial sample until the unconditional p-value is less than an assigned significance level. That is, until the null hypothesis is rejected. If the null hypothesis is rejected we assume there is a shift in the mean parameter over that sample. Then we calculate the posterior probability of each possible two-regime classification for that sample. The classification having the greatest posterior probability of a change point is selected as the posterior most likely change point for the mean (i.e., generalized maximum likelihood estimator[5]). The first detected change point for the mean is $\hat{\tau}_1(m)$. We repeat the above procedure with an initial sample beginning at the first data point following the previously detected change point (i.e., $\hat{\tau}_1(m) + 1$). When the null H_{01} is rejected, we again calculate the posterior probability of each data point over the sample period beginning at $t = \hat{\tau}_1(m) + 1$. The two-regime classificaiton having the greatest posterior probability of change point is selected as the second change point for the mean, $\hat{\tau}_2(m)$. This procedure is repeated until all data (T observations) are scanned. Then all change points for the mean have been sequentially detected. The sequence of mean parameters is $\mu_1, \ldots, \mu_{K_m}$. The maximum likelihood estimate of μ_k is calculated by using observations from $t = \hat{\tau}_{k-1}(m) + 1$ to $t = \hat{\tau}_k(m)$.

For the stationarity test of the variance parameter unconditionally on any mean parameter change, we perform the same procedure by obtaining the unconditional p-value for the null hypothesis

$$H_{02} : \rho = \sigma_2^2/\sigma_1^2 = 1$$

over a given test period, where σ_1^2 and σ_2^2 are also variance parameters of two implicit regimes. As in the detection procedure of change points for the mean, change points for the variance, $\tau_1(v), \tau_2(v), \ldots, \tau_{K_v-1}(v)$, are also sequentially estimated. The sequence of variance parameters are $\sigma_1^2, \ldots, \sigma_{K_v}^2$. The maximum likelihood estimate of σ_k^2 is calculated by using observations from $t = \hat{\tau}_{k-1}(v) + 1$ to $t = \hat{\tau}_k(v)$.

The sequential detection of change points methodology requires (1) A computational method for the unconditional p-values, and (2) the posterior distribution of a change point, τ, in each test period. We calculate the p-values based on the highest posterior density (HPD) interval. This is a Bayesian significance test with a diffuse prior (see in detail Lindley (1965), Box and Tiao (1973), and Kim (1991)).[6] This HPD interval test can be used when the information on target parameters is diffuse, while the traditional Bayes test using a posterior odds cannot. Since we have no information on the change of mean and variance parameters of stock returns a priori, the HPD interval test is appropriate.[7]

Based on Box and Tiao (1973), the unconditional p-value by the HPD interval test H_{01} over a testing period of sample size n is approximated as

$$p_{\delta=0} \approx \sum_{\tau} 2 \{1 - \Im_b(|t'(0)|)\} \ \pi(\tau|R), \tag{1}$$

where

$$t'(\delta) = (\delta - \hat{\delta}) \ / \ \{a^2(s_1^2/n_1 + s_2^2/n_2)\}^{1/2}$$

$$n_1 = \tau, \qquad n_2 = n - \tau$$

$$\hat{\delta} = \sum_{t=1}^{\tau} R_t/n_1 - \sum_{t=\tau+1}^{n} R_t/n_2 \equiv \bar{R}_1 - \bar{R}_2$$

$$s_1^2 = \sum_{t=1}^{\tau} (R_t - \bar{R}_1)^2/(n_1 - 1), \qquad s_2^2 = \sum_{t=\tau+1}^{n} (R_t - \bar{R}_2)^2/(n_2 - 1)$$

$$a^2 = ((b - 2)/b)f_1, \qquad b = 4 + f_1^2/f_2$$

$$f_1 = ((n_2 - 1)/(n_2 - 3)) \cos^2\phi + ((n_1 - 1)/(n_1 - 3))\sin^2\phi$$

$$f_2 = \frac{(n_2 - 1)^2}{(n_2 - 3)^2(n_2 - 5)} \cos^4\phi + \frac{(n_1 - 1)^2}{(n_1 - 3)^2(n_1 - 5)} \sin^4\phi$$

$$\cos^2\phi = (s_2^2/n_2)/(s_1^2/n_1 + s_2^2/n_2), \qquad \sin^2\phi = 1 - \cos^2\phi,$$

and $\mathfrak{I}_b(\cdot)$ is the cumulative density function of the standard Student t distribution with b degrees of freedom, and $\pi(\tau|R)$ is the posterior distribution of a change point τ. Conditional on a given change point τ, the inside bracket in equation (1) represents the stationarity test of mean parameter unconditional on any variance change. It computes the conditional p-value for H_{01} given the change point. Since we have no information on the change point, we assign every possible value of τ into the computation of the conditional p-value. Then the unconditional p-value for H_{01} is the weighted average of the conditional p-values for H_{01} given a change point τ with the probability weights determined by $\pi(\tau|R)$. In this approximation, each regime should have at least six observations; that is, $6 \leq \tau \leq n - 6$. However, to avoid any spurious detection of mean or variance parameter nonstationarity from too small a sample, we require at least seven data points in each regime.

The posterior probability of τ is determined by

$$\pi(\tau|R) = \pi(\tau)n_1 n_2 \Gamma((n_1 - 1)/2)\Gamma((n_2 - 1)/2)S_1^{(n_1-1)/2}S_2^{(n_2-1)/2} \tag{2}$$

where $S_1 = \Sigma_{t=1}^{\tau}(R_t - \bar{R}_1)^2$, $S_2 = \Sigma_{t=\tau+1}^{n}(R_t - \bar{R}_2)^2$, $\Gamma(\cdot)$ is a gamma function, and $\pi(\tau)$ is a prior distribution of τ. We can assign a uniform prior to τ because we have no prior information on τ, and hence, the term $\pi(\tau)$ has no impact in determining the posterior probability of τ.

The unconditional p-value for H_{02} by the HPD interval test for sample size n is given by

$$P_{\rho=1} = \sum_{\tau} 2\{1 - \Upsilon_{\tau-1,n-\tau-1}(F(1))\} \pi(\tau|R), \tag{3}$$

where $\Upsilon_{\tau-1,n-\tau-1}(\cdot)$ is the cumulative density function of an F distribution with $(\tau - 1, n - \tau - 1)$ degrees of freedom, and

$$F(\rho) = \max\{\rho(\hat{\sigma}_1^2/\hat{\sigma}_2^2), \rho(\hat{\sigma}_2^2/\hat{\sigma}_1^2)\}.$$

Note that $F(1)$ is the conditional F-test statistic for H_{02} given a change point τ. Therefore, the unconditional p-value test for variance parameter stationarity is also the weighted average of the conditional p-values based on an F-test given τ (i.e., the inside bracket in equation (3)) with the probability weights determined by the posterior distribution of τ in equation (2). As in the mean stationarity test, we require at least seven data points in each regime. If the unconditional p-value for H_{02} is less than an assigned significance level, we also calculate the posterior probability (equation (2)) of each possible two-regime classification. Then we select the two-regime classification having the greatest posterior probability as the posteriori most likely change point for the variance parameter.

The derivation of the unconditional p-values fo H_{01} and H_{02} and the posterior distribution of τ are described in the appendix.

2.5. Sequential mixture model simulation

This simulation provides a justification for the sequential mixture estimation prodedure and the selection of a 1% significance level detection criteria. The number of observations in the time series of the simulated data is 100 in table 3 and 400 in table 4. In every case, 100 replications are used throughout tables 3 and 4. THe simulated data is generated from a normal distribution with a mean of 0.05% and standard deviation of 1%. Then two change points for the variance parameter (three regimes) are randomly chosen over the time series. When the predetermined ratio of standard deviations is set to be 2, the standard deviation of the second regime is twice that of the first regime. and the standard deviation of the third regime is half that of the second regime. The same procedure is sued for the predetermined ratios of 3, 4, and 5. The mean parameter remains fixed over the sample period.

Table 3 presents the estimation results for comparison of the simultaneous and the sequential detection methods. Panel A results are from the simultaneous method assuming that the number of regimes is unknown. For each potential number of regimes from 1 to 5, we compute the Schwarz (1978) values (large sample posterior probability) for all possible classifications of the data into regimes. The set of data points with the maximum Schwarz value is selected as the estimate of the change points for variance. Panel B presents the estimation results by the simultaneous method by assuming that the number of regimes is exactly known. Since we do not know the number of regimes a priori, panel B values represent benchmarks for comparison. The results in panel A indicate a substantial amount of overdetection. When the predetermined ratio of standard deviations is 1, the true model has only one regime and no change points. The simultaneous method, however, estimates change points even though none exist. For true models with three regimes and a predetermined standard deviation ratio of 2, there is some underdetection (.04 or 4%), but mostly overdetection (.65 or 65% for 4 and 5+ regimes). The overdetection increases as the predetermined ratio of standard deviations increase.

Panels C, D, and E are the estimation results by the sequential detection method (HPD interval test) at the 1, 5, and 10% significance levels, respectively. At the 1% significance level there is a small amount of overdetection throughout the predetermined ratios. There is, however, some undertection. Particularly at the predetermined ratios of 2 and 3, but declining at ratios of 4 and 5. As the significance level for detection increases to 5% and 10% in panels D and E, the overdetection increases and Schwarz values decline.

Table 3. Simulation results: Estimation of unknown change points for variance by the simultaneous and sequential detection methods when two change points (three regimes) are randomly given over the sample size of 100 (100 replications).

Ratio of Standard Deviations[1] ($\rho^{1/2}$)	Frequency of Estimated Number of Regimes						Average Exact Schwarz Value (Standard Error)	Average Estimated Schwarz Value[2] (Standard Error)
	1	2	3	4	5+	Ave		
Panel A: By Simultaneous Detection Method When the Number of Regimes Is Unknown								
1	0.00	0.07	0.31	0.32	0.27	3.70	314.8 (0.64)	306.7 (5.47)
2	0.00	0.04	0.31	0.28	0.37	3.98	288.2 (1.52)	291.4 (1.52)
3	0.00	0.00	0.33	0.21	0.46	4.13	272.5 (2.12)	276.6 (2.27)
4	0.00	0.00	0.28	0.23	0.49	4.21	263.1 (2.40)	265.9 (2.83)
5	0.00	0.00	0.27	0.24	0.49	4.25	253.4 (3.04)	256.3 (3.28)
Panel B: By Simultaneous Detection Method When the Number of Regimes Is Known								
1	1.	0.	0.	0.	0.	1.00	314.8 (0.64)	314.8 (0.64)
2	0.	0.	1.	0.	0.	3.00	288.2 (1.52)	290.7 (1.52)
3	0.	0.	1.	0.	0.	3.00	272.5 (2.12)	275.6 (2.26)
4	0.	0.	1.	0.	0.	3.00	263.1 (2.40)	264.8 (2.81)
5	0.	0.	1.	0.	0.	3.00	253.4 (3.04)	256.1 (3.26)
Panel C: By Sequential Detection Method at the 1% Significance Level for the HPD Test[3]								
1	0.87	0.09	0.04	0.00	0.00	1.17	314.8 (0.64)	315.1 (0.67)
2	0.24	0.24	0.49	0.03	0.00	2.31	288.2 (1.52)	288.5 (1.55)
3	0.00	0.21	0.76	0.03	0.00	2.82	272.5 (2.12)	271.7 (2.08)
4	0.00	0.07	0.88	0.05	0.00	2.98	263.1 (2.40)	261.5 (2.33)
5	0.00	0.08	0.90	0.02	0.00	2.97	253.4 (3.04)	251.5 (2.96)
Panel D: By Sequential Detection Method at the 5% Significance Level for the HPD Test								
1	0.62	0.19	0.18	0.01	0.00	1.58	314.8 (0.64)	314.6 (0.74)
2	0.10	0.18	0.56	0.13	0.03	2.81	288.2 (1.52)	288.8 (1.53)
3	0.00	0.08	0.76	0.11	0.05	3.13	272.5 (2.12)	272.1 (2.11)
4	0.00	0.02	0.74	0.22	0.02	3.24	263.1 (2.40)	261.1 (2.33)
5	0.00	0.04	0.76	0.18	0.02	3.19	253.4 (3.04)	251.2 (2.93)
Panel E: By Sequential Detection Method at the 10% Significance Level for the HPD Test								
1	0.40	0.25	0.27	0.06	0.02	2.05	314.8 (0.64)	314.0 (0.73)
2	0.06	0.14	0.52	0.22	0.06	3.08	288.2 (1.52)	288.6 (1.54)
3	0.00	0.05	0.67	0.19	0.09	3.32	272.5 (2.12)	271.9 (2.12)
4	0.00	0.03	0.57	0.33	0.07	3.44	263.1 (2.40)	260.7 (2.33)
5	0.00	0.04	0.67	0.24	0.05	3.32	253.4 (3.04)	250.1 (2.94)

[1]This indicates the ratio of standard deviations between two adjacent regimes. When the ratio is one, there is no change point in the variance parameter (i.e., one regime).

[2]When two change points are estimated by random selection, the average Schwarz values are 314.8, 281.7, 258.2, 241.3, and 224.6 for $\rho^{1/2} = 1$, 2, 3, 4, and 5, respectively.

[3]The minimum sample size is seven in each regime, and the initial testing sample size is 30.

Table 4. Simulation results: Estimation of unknown change points for variance by the simultaneous and sequential detection methods when two change points (three regimes) are randomly given over the sample size of 400 (100 replications).

Ratio of Standard Deviations[1] ($\rho^{1/2}$)	Frequency of Estimated Number of Regimes						Average Exact Schwarz Value (Standard Error)	Average Estimated Schwarz Value[2] (Standard Error)
	1	2	3	4	5+	Ave		
Panel A: By Simultaneous Detection Method When the Number of Regimes Is Known								
1	1.	0.	0.	0.	0.	1.00	1,271.4 (1.49)	1,271.4 (1.49)
2	0.	0.	1.	0.	0.	3.00	1,167.8 (6.02)	1,169.6 (6.06)
3	0.	0.	1.	0.	0.	3.00	1,127.5 (9.17)	1,128.7 (9.52)
4	0.	0.	1.	0.	0.	3.00	1,082.6 (11.50)	1,084.3 (12.01)
5	0.	0.	1.	0.	0.	3.00	1,073.6 (13.31)	1,074.9 (13.97)
Panel B: By Sequential Detection Method at the 1% Significance Level for the HPD Test[3]								
1	0.75	0.10	0.14	0.01	0.00	1.41	1,271.4 (1.49)	1,270.7 (1.54)
2	0.01	0.14	0.67	0.12	0.06	3.08	1,167.8 (6.02)	1,168.0 (5.95)
3	0.00	0.03	0.65	0.12	0.20	3.52	1,127.5 (9.17)	1,126.9 (0.08)
4	0.00	0.01	0.65	0.17	0.17	3.54	1,082.6 (11.50)	1,078.7 (11.64)
5	0.00	0.01	0.71	0.16	0.12	3.42	1,073.6 (13.31)	1,069.3 (13.40)
Panel C: By Sequential Detection Method at the 5% Significance Level for the HPD Test								
1	0.25	0.17	0.28	0.08	0.21	3.05	1,271.4 (1.49)	1,266.9 (1.68)
2	0.00	0.02	0.27	0.22	0.49	4.70	1,167.8 (6.02)	1,165.0 (5.95)
3	0.00	0.02	0.27	0.19	0.52	4.90	1,127.5 (9.17)	1,123.4 (9.15)
4	0.00	0.01	0.18	0.11	0.70	5.49	1,082.6 (11.50)	1,074.1 (11.68)
5	0.00	0.00	0.18	0.23	0.59	5.28	1,073.6 (13.31)	1,066.4 (13.32)
Panel D: By Sequential Detection Method at the 10% Significance Level for the HPD Test								
1	0.05	0.04	0.18	0.22	0.51	5.07	1,271.4 (1.49)	1,262.6 (1.75)
2	0.00	0.02	0.07	0.05	0.31	6.17	1,167.8 (6.02)	1,162.6 (5.95)
3	0.00	0.01	0.13	0.09	0.87	6.79	1,127.5 (9.17)	1,120.0 (9.15)
4	0.00	0.00	0.05	0.04	0.91	7.27	1,082.6 (11.50)	1,070.5 (11.70)
5	0.00	0.00	0.04	0.07	0.89	7.15	1,073.6 (13.31)	1,061.9 (13.24)

[1]This indicates the ratio of standard deviations between two adjacent regimes. When the ratio is one, there is no change point in the variance parameter (i.e., one regime).

[2]When two change points are estimated by random selection, the average Schwarz values are 1,271.4, 1,144.8, 1,071.0, 995.3, and 961.8 for $\rho^{1/2}$ = 1, 2, 3, 4, and 5, respectively.

[3]The minimum sample size is seven in each regime, and the initial testing sample size is 30.

In table 4, the same procedures are used for a larger time series of 400 observations. Excluded from table 4 is the simultaneous detection method when the number of regimes is unknown. This is the situation that becomes computationally difficult as the number of observations and/or number of regimes increase. We do, however, wish to observe the properties of the sequential detection procedure as the sample size increases. All of the comparisons between the panels are essentially the same as the 100 observation case. The major difference between tables 3 and 4 is a general decrease in underdetection and an increase in overdetection by the sequential detection method. Given the best results are found with the 1% significance level, this criteria is used on actual data throughout this paper.

3. The competing models

The competing models for the distribution of daily stock returns are the Student t distribution (s), the Poisson jump-diffusion process (jp), the generalized discrete mixture of normal distribution (gm), and the sequential mixture of normal distribution (sm) models.

3.1. Student t (s)

Blattberg and Gonedes (1974) assume that the variance parameter of a normally distributed stock return is a drawing from an inverted gamma distribution. Then the resulting posterior distribution is the Student model. The Student t probability density function with location parameter μ_s, scale parameter H_s, and degrees of freedom ν_s is represented by

$$f(R_t \mid \Theta_s) = \frac{\Gamma[(\nu_s + 1)/2] \sqrt{H_s}}{\Gamma[\nu_s/2] \sqrt{\pi \nu_s}} \left\{ 1 + \frac{H_s}{\nu_s} (R_t - \mu_s)^2 \right\}^{-(\nu_s+1)/2},$$

where $\Theta_s = \{\mu_s, H_s, \nu_s; -\infty < \mu_s < \infty, H_s > 0, \nu_s > 0\}$.

The estimates that maximize the log-likelihood function for the Student t model and our data sample are contained in table 5. The student distribution approaches the normal as the degrees-of-freedom parameter gets large. In order to explain the observed kurtosis relative to the stationary normal exhibited in table 1, the degrees-of-freedom parameter for the student model should be in the range $2 < \nu_s < 10$. This is confirmed for our sample in table 5. Furthermore, the scale parameter is inversely related to the variance of the distribution. Therefore, the indexes have considerably larger scale estimates and lower variances than the individual stocks.

3.2. Poisson jump-diffusion (jp)

The Press (1968) Poisson jump-diffusion process for the distribution of stock returns consists of a geometric Brownian motion components and an independent compound Poisson process with normally distributed jump amplitudes. This Poisson jump-diffusion process can be expressed as follows:

$$\log \left(\frac{S(t)}{S(t-1)} \right) = \left(\alpha - \frac{1}{2} \sigma^2 \right) t + \sigma Z(t) + \sum_{n=1}^{N(t)} J_n,$$

where $S(t)$ = the security price at time t; $Z(t)$ = a standard Brownian motion; $N(t)$ = a Poisson counting process with parameter $\lambda > 0$; J_n = a normal random variate with mean μ_J and variance σ_J^2 representing the logarithm of 1 plus the percentage change in security price caused by the nth jump; α = the instantaneous conditional expected rate of return per unit time for the Brownian motion part of the process; and σ^2 = the instantaneous conditional variance of α.

Table 5. Student *t* distribution estimators.

I.D. No.	$\mu_s(\times 10^3)$ (*t*-Statistic)	$H_s(\times 10^{-3})$ (*t*-Statistic)	$\nu_s(\times 1/2)$ (*t*-Statistic)	I.D. No.	$\mu_s(\times 10^3)$ (*t*-Statistic)	$H_s(\times 10^{-3})$ (*t*-Statistic)	$\nu_s(\times 1/2)$ (*t*-Statistic)
1	0.0410 (0.11)	7.9943 (14.58)	1.6213 (10.67)	18	0.1349 (0.77)	6.2042 (34.32)	2.3922 (17.53)
2	0.1007 (0.56)	5.7861 (33.35)	2.3187 (17.42)	19	0.6045 (2.78)	4.8328 (31.01)	1.9499 (19.11)
3	0.1843 (0.61)	4.2990 (23.61)	2.3686 (12.15)	20	0.1659 (0.93)	8.0051 (29.21)	2.1363 (16.51)
4	−0.0645 (−0.58)	16.7063 (30.67)	1.6319 (22.00)	21	0.4549 (2.96)	7.8565 (33.62)	2.5737 (15.91)
5	−0.5373 (−2.76)	5.3941 (32.05)	1.7097 (22.15)	22	0.2124 (1.46)	8.9850 (33.580)	2.2400 (18.16)
6	0.0265 (0.12)	3.8087 (32.85)	2.3139 (17.87)	23	0.6158 (3.77)	7.1148 (33.04)	2.2757 (17.53)
7	0.3073 (1.85)	7.0255 (33.07)	2.0124 (19.73)	24	0.2498 (1.97)	12.2415 (32.46)	1.9656 (19.73)
8	0.2752 (1.76)	8.0508 (31.72)	1.9064 (19.76)	25	−0.0735 (−0.49)	8.9714 (32.27)	1.8607 (20.63)
9	0.4081 (2.77)	9.1235 (33.07)	1.9010 (20.82)	26	0.1778 (1.20)	9.1752 (32.58)	1.7887 (21.65)
10	0.4478 (2.14)	4.4931 (32.69)	2.0112 (19.47)	27	−0.1575 (−1.00)	8.1409 (32.49)	1.7985 (21.42)
11	−0.0159 (−0.11)	9.2386 (32.26)	2.0974 (18.41)	28	0.2092 (1.11)	5.2701 (33.350)	2.3712 (17.05)
12	0.0692 (0.43)	7.4593 (34.10)	2.2100 (18.76)	29	0.0318 (0.17)	5.5234 (33.30)	1.9671 (20.33)
13	0.4800 (3.69)	11.0978 (34.19)	2.4144 (17.30)	30	−0.1054 (−0.59)	6.1116 (32.45)	2.0483 (18.97)
14	0.2677 (1.78)	8.4303 (33.80)	2.2698 (18.08)	S&P	0.3174 (3.70)	26.6409 (32.43)	1.9459 (19.90)
15	0.1054 (0.71)	8.7256 (33.66)	2.2945 (17.81)	EW	1.0582 (15.01)	42.1660 (31.75)	1.54 14 (24.10)
16	−0.1183 (−0.67)	6.1454 (34.46)	2.2822 (18.43)	VW	0.5471 (6.75)	30.0633 (32.01)	1.8835 (20.20)
17	0.1858 (1.29)	8.8900 (34.93)	2.6118 (16.49)				

Then, the probability density function of a security return is

$$f(R_t \mid \Theta_{jp}) = \sum_{n=1}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \cdot \phi(R_t \mid \mu + n\mu_J, \ \sigma^2 + n\sigma_J^2),$$

where $\mu = \alpha - 1/2\sigma^2$, $\Theta_{jp} = \{\mu, \sigma^2, \lambda, \mu_J, \sigma_J^2\}$, and $\phi(R_t \mid a, b)$ denotes a normal density function with mean a and variance b.

The estimates that maximize the log-likelihood function for the Poisson jump-diffusion process and our sample data are contained in table 6. For all of the individual securities and indexes that converged to an optimum, all but one security had an estimate of either λ or σ_f^2 that is statistically significant. Therefore, this infinite mixture of normal distributions model is also capable of explaining the observed kurtosis in table 1.

3.3. Generalized discrete mixture of normal distrubitons (gm)

A generalized discrete mixture of normal distributions alternative for daily stock returns has been proposed and tested in Kon (1984). Each return observation is viewed as a drawing from one of a finite number of normal distributions with some probability. The observations are not necessarily consecutive. Therefore, the probability density function for L regimes is

$$f(R_t \mid \Theta_{gm}) = \sum_{j=1}^{L} \lambda_j \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{ -\frac{1}{2\sigma_j^2} (R_t - \mu_j)^2 \right\},$$

where the λ_js are the mixing probabilities and the parameter vector is $\Theta_{gm} = \{\mu_j, \sigma_j^2, \lambda_j\}_{j=1,\ldots,L}$. A comparison of the stationary normal distribution with a mixture of two normal distributions can be made with the statistic: minus 2 times the log-likelihood ratio of the models. This statistic has an asymptotically chi-square distribution with 3 degrees of freedom. At the .01 probability level, the statistic must exceed 11.3. The actual values ranged from 423.26 to 1,854.38. Hence, mixtgures of three, four, and five normal distributions can explain the observed kutosis and skewness in table 1.

3.4. Sequential mixture (sm)

The sequential mixture of normal distributions model is conditioned on the estimated change points, $\hat{\tau}_1(m), \ldots, \hat{\tau}_{k_m-1}(m), \hat{\tau}_1(v), \ldots, \hat{\tau}_{k_v-1}(v))$, where K_m and K_v are the total number of change points for the mean and for the variance, respectively. Thus, the parameter set of this model is $\Theta_{sm} = (\mu_1, \ldots, \mu_{K_m}, \sigma_1^2, \sigma_{K_v}^2)$. Conditional on the estimated change points, the maximum likelihood estimate (MLE) of μ_k is the sample mean ($\hat{\mu}_k$) calculated by using observations from $t = \hat{\tau}_{k-1}(m) + 1$ to $t = \hat{\tau}_k(m)$, and the MLE of σ_k^2 is also the sample variance ($\hat{\sigma}_k^2$) calculated by using observations from $t = \hat{\tau}_{k-1}(v) + 1$ to $t = \hat{\tau}_k(v)$. Let μ_t and σ_t^2 be the mean and variance of a stock return at a specific time t, R_t. Then the MLE of μ_t, $\hat{\mu}_t$, equals $\hat{\mu}_k$ for $t \in [\hat{\tau}_{k-1}(m) + 1, \hat{\tau}_k(m)]$, and the MLE of σ_t^2, $\hat{\sigma}_t^2$, equals $\hat{\sigma}_k^2$ for $t \in [\hat{\tau}_{k-1}(v) + 1, \hat{\tau}_k(v)]$. Therefore, the maximum likelihood value of Θ_m is calculated by

$$L(\hat{\Theta}_{sm} \mid R) = (2\pi)^{-T/2} \prod_{t=1}^{T} \left\{ (\hat{\sigma}_t^2)^{-1/2} \exp\left\{ -\frac{1}{2\hat{\sigma}_t^2} (R_t - \hat{\mu}_t)^2 \right\} \right\}.$$

Table 6. Poisson jump diffusion process estimators.

I.D. No.	$\mu(\times 10^3)$	$\sigma^2(\times 10^3)$	$\mu_J(\times 10^3)$	$\sigma_J^2(\times 10^3)$	λ	I.D. No.	$\mu(\times 10^3)$	$\sigma^2(\times 10^3)$	$\mu_J(\times 10^3)$	$\sigma_J^2(\times 10^3)$	λ
1	0.2924	0.1873	−2.0929	5.6731	0.0379	18	−0.3443	0.1470	3.7928	0.5004	0.2432
	(0.93)	(4.21)	(−1.54)	(1.98)	(5.21)		(−1.55)	(1.39)	(1.90)	(4.90)	(3.11)
2	0.7151	0.3225	−0.4481	0.0031	0.2346	19	0.2664	0.1936	3.3957	0.8551	0.2450
	(1.32)	(2.09)	(−1.81)	(1.90)	(3.21)		(0.90)	(2.54)	(2.11)	(3.46)	(1.67)
3	−0.1079	0.2195	3.3885	0.7402	0.2375	20	−0.2407	0.1173	3.9562	0.4682	0.2351
	(−0.78)	(2.12)	(2.67)	(6.32)	(4.66)		(−0.540)	(1.39)	(1.59)	(4.19)	(2.56)
4	−0.6991	0.0330	1.8206	0.1570	0.6211	21	N.C.	N.C.	N.C.	N.C.	N.C.
	(−1.09)	(3.44)	(0.87)	(4.20)	(2.00)		(N.C.)	(N.C.)	(N.C.)	(N.C.)	(N.C.)
5	−1.0838	0.1693	5.3333	0.8495	0.2687	22	−0.0889	0.1025	2.6066	0.3786	0.2406
	(−2.31)	(12.09)	(2.45)	(1.09)	(2.43)		(−0.06)	(1.23)	(1.09)	(2.45)	(3.17)
6	N.C.	N.C.	N.C.	N.C.	N.C.	23	0.2679	0.1296	2.9745	0.4454	0.2461
	(N.C.)	(N.C.)	(N.C.)	(N.C.)	(N.C.)		(0.93)	(2.05)	(1.67)	(4.55)	(2.32)
7	0.1818	0.1303	1.3059	0.5399	0.2568	24	−0.1758	0.0547	1.4651	0.1999	0.5061
	(0.98)	(1.09)	(0.67)	(2.01)	(3.55)		(−1.02)	(1.00)	(0.56)	(1.34)	(4.56)
8	−0.2733	0.0663	1.2727	0.2436	0.7023	25	−0.5564	0.0787	2.0060	0.3076	0.4611
	(−0.70)	(0.23)	(1.34)	(3.54)	(5.21)		(−1.54)	(0.90)	(1.43)	(3.20)	(2.78)
9	0.1633	0.1028	2.4842	0.4794	0.2409	26	−0.0464	0.1047	2.5488	0.5239	0.2333
	(0.88)	(2.44)	(2.04)	(2.60)	(1.87)		(−0.06)	(1.76)	(1.22)	(4.29)	(2.40)
10	−0.0044	0.2054	4.2768	0.8428	0.2494	27	−0.7411	0.1019	3.5945	0.4403	0.3369
	(−0.12)	(3.44)	(2.08)	(1.98)	(1.30)		(−1.78)	(2.10)	(1.45)	(4.32)	(3.20)
11	−0.8712	0.0611	1.9698	0.1981	0.6632	28	−0.6680	0.1184	2.1099	0.3123	0.6251
	(−1.87)	(0.59)	(1.56)	(1.430)	(4.65)		(−1.32)	(3.43)	(0.98)	(3.21)	(5.09)
12	−0.2668	0.1268	3.3490	0.4774	0.2356	29	−0.4779	0.1664	0.4644	0.7038	0.2596
	(−1.33)	(2.11)	(2.23)	(1.60)	(0.71)		(−0.93)	(2.78)	(1.090)	(5.21)	(3.78)
13	0.3448	0.0837	1.3184	0.2923	0.2317	30	−1.4069	0.0978	3.4134	0.3290	0.5968
	(1.54)	(0.49)	(1.86)	(3.48)	(1.95)		(−1.89)	(2.13)	(1.09)	(3.21)	(4.38)
14	0.0063	0.1098	2.3425	0.3993	0.2384	S&P	0.2804	0.0360	0.1011	0.1615	0.2295
	(0.10)	(1.76)	(1.66)	(3.25)	(2.09)		(2.90)	(1.87)	(2.90)	(1.56)	(3.45)
15	−0.1989	0.1064	2.7624	0.3767	0.2387	EW	1.2644	0.0223	−2.2028	0.1389	0.2490
	(−1.20)	(1.90)	(2.03)	(4.19)	(1.45)		(4.59)	(2.33)	(−3.29)	(4.56)	(3.29)
16	−0.5533	0.1497	4.1086	0.5473	0.2405	VW	0.5680	0.0332	−0.5222	0.1590	0.2069
	(−1.98)	(1.22)	(1.90)	(4.90)	(2.44)		(2.78)	(1.98)	(−3.59)	(4.21)	(2.21)
17	−0.1252	0.1019	2.4903	0.3048	0.2545						
	(−0.490)	(1.42)	(1.09)	(4.02)	(2.65)						

Note: N.C. indicates no convergence.

Table 7 provides the number of change points that were estimated for each stock and index time-series at the 1% significance level. The large number of change points for the variance in each time series indicates that the sequential mixture model of normal distributions is also consistent with the observed kurtosis in table 1.

There is no general rule for the comparison of number of change points detected and portfolio size. The fact that there are more change points detected for the variance of the equally weighted index than change points for the variance of individual securities is consistent with nonsynchronous calendar changes across securities. For example, if two stocks had one change point (two regimes) each at different dates, then a portfolio of these two stocks will have two change points (three regimes). Thus, the process of aggregation may result in a portfolio with more nonstationarity of variance over time than their components

Table 7. Number of detected change points by the HPD interval test on the mean and variance of daily stock returns at the 1% significance level.

I.D. No.	No. of Mean Change Points	No. of Variance Change Points
1	1	15
2	5	65
3	0	26
4	1	80
5	2	72
6	5	64
7	5	81
8	2	79
9	1	61
10	9	65
11	5	83
12	1	66
13	1	67
14	2	71
15	0	55
16	3	69
17	2	65
18	3	81
19	4	47
20	4	57
21	4	71
22	1	62
23	6	81
24	3	65
25	2	71
26	2	81
27	2	76
28	4	69
29	5	86
30	0	68
31	10	73
32	67	97
33	12	78

(individual stocks).[8] Aggregation may also result in detecting fewer change points if the changes of component variances are offsetting or if the investment weights create a zero market beta portfolio. Cross-sectionally, we expect to observe synchronous changes in the market component of variance as well as nonsynchronous changes in the firm-specific component of variance.

More change points for the mean of the index than change points for the mean of individual securities is also due to the sensitivity of the test to the magnitude of the variance parameter. That is, if variance is small, the change in the mean can be relatively easily detected. However, if variance is large for the same magnitude of change in the mean, the change in the mean is hard to detect. It is well known that the variance of portfolios (indexes) is much smaller than that of individual securities (i.e., see table 1).
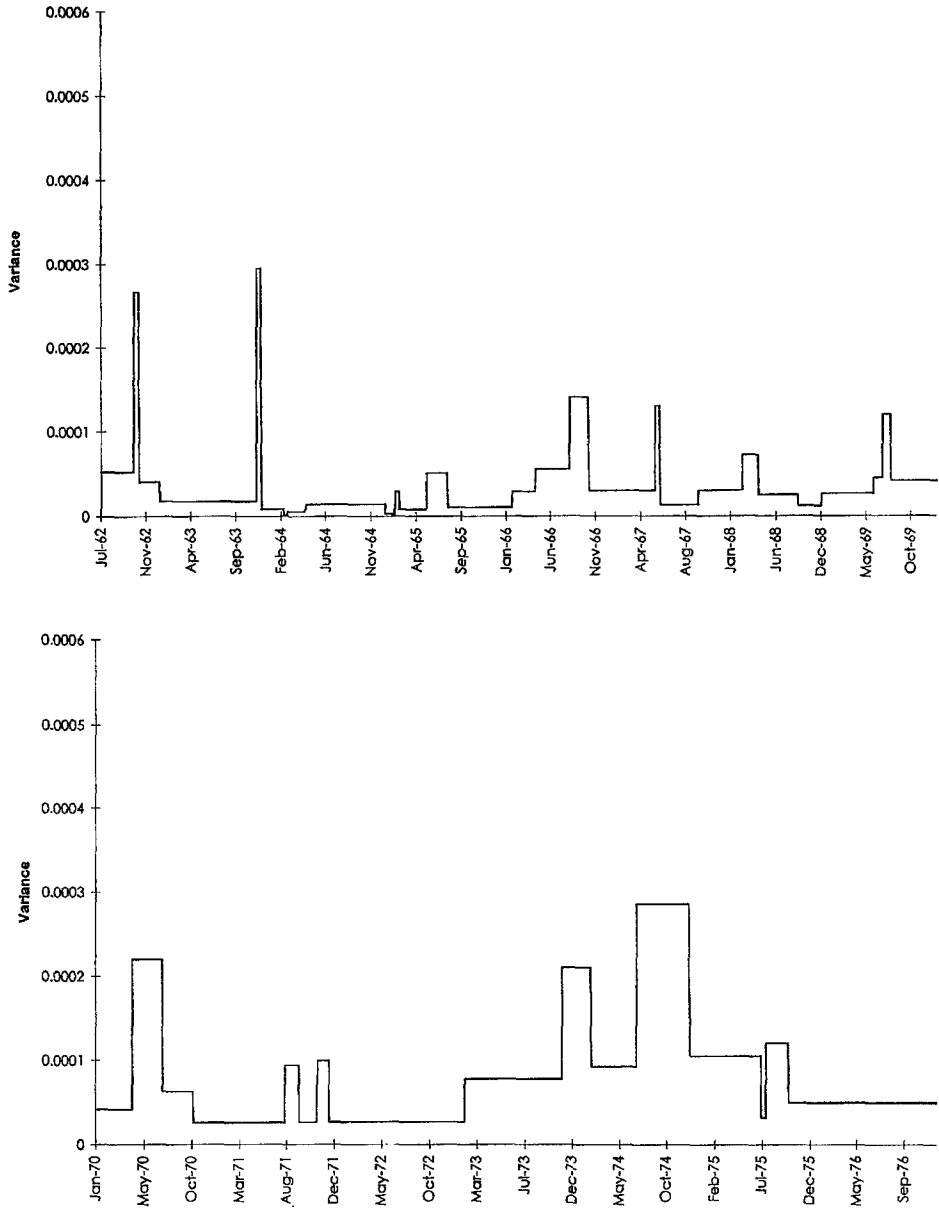
Figure 1. Sequential mixture variances for the S&P 500.

Figure 1 displays the sequential mixture variance estimates for the S&P 500 stock index. The results are consistent with our intuition. Without a priori knowledge, regimes with high variance estimates include the Cuban missile crisis, the assassination of President John F. Kennedy, the invasion of Cambodia, President Nixon's resignation, the stock market crash
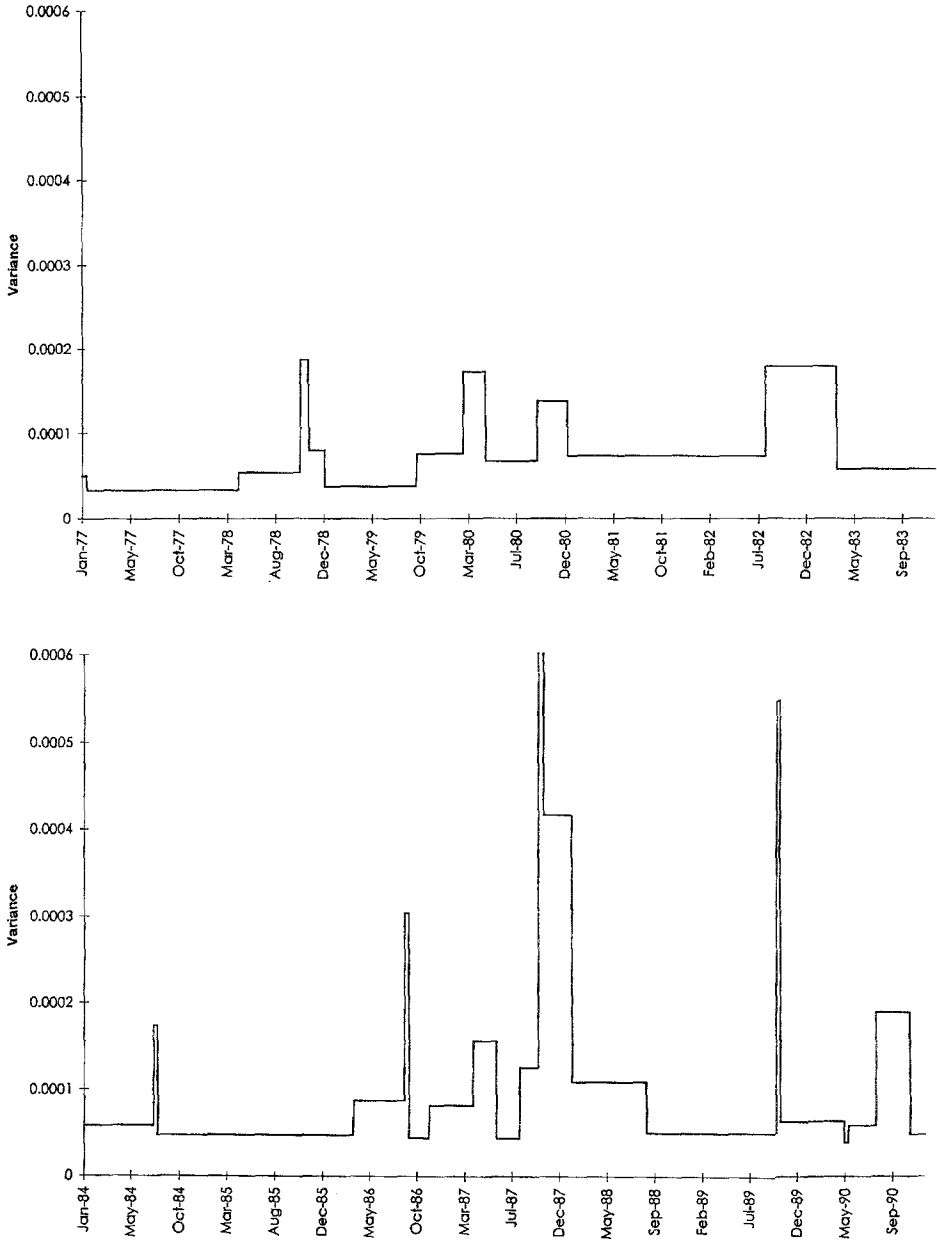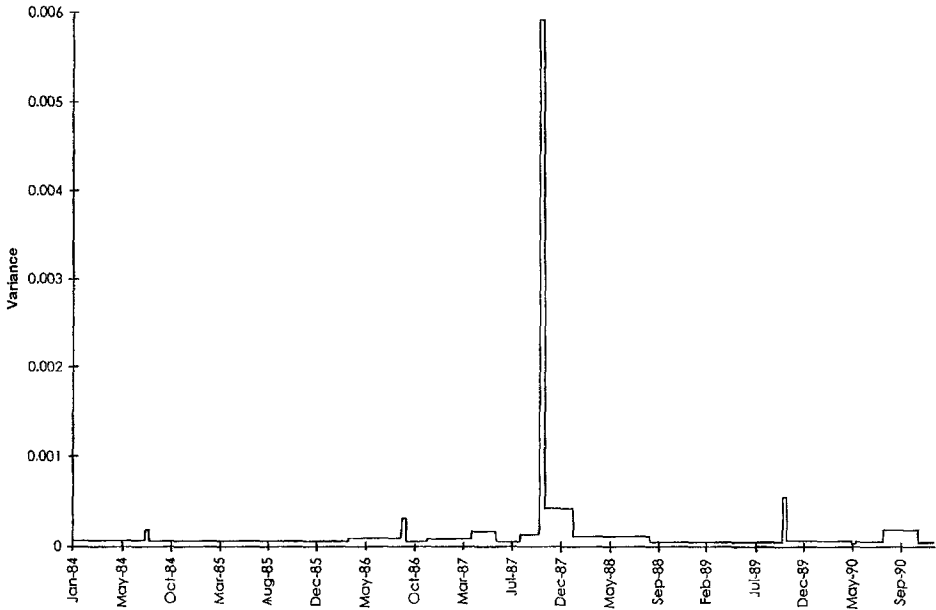
Figure 1. Continued.

Figure 1. Continued.

of October 19, 1987, the minicrash of October 13, 1989, and the Iraqi invasion of Kuwait. The procedure appears to be capable of detecting the major market volatility changes.

The logical next step is to determine which of these alternatives represents the best description of daily stock returns.

4. Model of stock returns—A comparison

The log-likelihood ratios of each competing model, against the stationary normal distribution model, are presented in table 8 for daily return data from July 2, 1962 to December 31, 1990. Note that all of the log-likelihood ratios in table 8 are positive. Therefore, each competing model might have some ability to capture the observed kurtosis in the unconditional distribution of stock returns. The quarterly partition is offered as the most likely of the partitions of the data from table 2. For 29 of the 30 stocks and all three indexes, the quarterly partition log-likelihoods are greater than the Student t, Poisson jump process, and the generalized mixture of normals alternatives. The log-likelihood dominance of the sequential mixture indicates that estimating sequential change points for parameter shifts may be superior to the assumption that parameter shifts can take place only at the beginning of each quarter. For the 28½-year time period, the quarterly partition represents 113 change points. The number of change points in table 7 for the sequential mixture model is considerably less for each stock and index.

The competing models in this study are not only nonnested, but also contain very different numbers of process parameters. Higher-order models have more process parameters

Table 8. Log-likelihood ratio of each model against the normal distribution model with daily returns from July 2, 1962 to December 31, 1990.

Firm's I.D. No.	SM	Student t	Poisson Jump	Two Normals	Three Normals	Four Normals	Five Normals	Quarterly Partition
1	488.86	376.56	353.88	309.25	315.88	327.77	334.04	362.32
2	765.49	387.25	290.87	347.38	400.46	412.58	417.37	626.01
3	422.21	239.48	220.12	211.63	250.96	N.C.	256.17	364.14
4	1,573.27	778.85	772.93	714.55	816.25	830.01	N.C.	1,250.30
5	1,286.78	764.21	738.28	703.82	784.67	798.82	N.C.	1,053.08
6	700.33	353.07	N.C.	351.71	384.04	N.C.	391.24	591.36
7	835.24	514.59	491.72	470.89	516.28	525.62	N.C.	628.61
8	1,043.28	447.74	450.12	433.92	462.86	N.C.	N.C.	847.17
9	1,110.31	681.53	650.70	616.79	682.08	N.C.	N.C.	957.76
10	959.01	529.89	505.09	476.97	548.93	N.C.	N.C.	830.30
11	943.16	399.26	411.15	379.79	426.01	434.10	442.70	759.20
12	1,115.79	642.29	602.59	556.13	N.C.	N.C.	571.72	943.88
13	971.22	524.59	486.63	454.19	N.C.	N.C.	N.C.	819.57
14	889.66	386.51	378.75	366.23	389.15	397.33	398.19	776.99
15	893.96	443.05	442.57	395.29	450.51	459.35	459.77	754.46
16	897.32	489.58	479.55	410.57	435.46	N.C.	512.68	705.48
17	852.45	396.63	378.20	350.78	N.C.	N.C.	N.C.	700.43
18	792.41	436.11	418.03	355.36	449.35	N.C.	N.C.	583.33
19	960.44	473.70	460.51	451.91	476.76	482.22	483.87	864.16
20	970.58	616.96	581.09	587.22	591.01	N.C.	N.C.	835.60
21	569.54	268.79	N.C.	251.11	N.C.	288.00	N.C.	435.36
22	830.93	497.73	470.54	461.15	506.92	N.C.	N.C.	725.91
23	730.53	321.88	317.22	N.C.	337.40	344.50	345.58	567.92
24	1,169.54	730.73	690.64	682.48	683.38	N.C.	719.63	1,039.25
25	1,150.96	623.50	605.41	530.30	623.60	644.54	645.20	955.63
26	1,343.61	637.12	607.05	N.C.	646.93	650.29	N.C.	1,095.74
27	1,241.33	680.66	655.95	628.78	694.98	708.43	711.56	1,008.84
28	668.19	307.63	307.75	294.73	321.71	325.94	N.C.	539.22
29	1,062.11	707.20	673.61	580.63	719.07	728.94	N.C.	805.29
30	745.42	453.66	471.08	430.94	488.59	N.C.	496.69	627.24
S&P	1,591.49	740.26	681.34	676.95	730.07	N.C.	N.C.	1,367.46
EW	2,171.17	1,013.48	978.45	927.19	1,032.95	1,047.28	1,048.25	1,600.39
VW	1,569.97	719.63	659.69	656.92	694.24	696.89	N.C.	1,326.88

Note: N.C. indicates no convergence.

[1]Sequential mixture of normal distributions.

and may have higher likelihood function values solely due to greater dimension. In order to control for this potential selection bias, we calculate the Schwarz (1978) criterion. This criterion provides guidance for selecting a model from a number of models with different parameter space by finding its large sample Bayesian solution. The Schwarz criterion is defined as

$$SC = \log(\text{maximum likelihood}) - \frac{1}{2} p \, \log(T),$$

where p is the number of estimated parameters. In the large sample limit the leading term in the Bayesian solution is the maximum likelihood estimator. The latter term in the Schwarz criterion represents the penalty for models of higher dimension (p). Thus, choose as most likely the model with the greatest SC value.

Table 9 contains the SC value of each competing model relative to the stationary normal distribution alternative. These are the asymptotic log posterior odds of each model relative to the independently and identically normally distributed alternative. It is quite striking to see that the quarterly partition alternative has negative relative SC values for 22 of 30 stocks. Clearly, the penalty for the increased number of parameters more than offsets any benefit of additional model fitting of the data. This evidence strongly contradicts the results of the log-likelihood ratios alone. Rather than ranking second best in table 8, the relative SC values in table 9 indicate that the quarterly partition actually ranks last among the competing models.

The sequential mixture of normal distributions model, however, has the highest relative SC values of all alternatives for 25 of the 30 stocks and all three indexes. It is evident that the descriptive value of this high dimension model is worthwhile. Hence, the benefit must come from the precise estimation of the change points for parameter nonstationarity.

It is also interesting to note that the Student t model ranks second in the sample of relative SC values. This model has the highest value for four stocks and is second highest for 22 of the 30 stocks and all three indexes. With only three parameters in the Student t model, the generalized mixture of normal distributions alternatives did not have enough additional descriptive ability in table 8 to offset the required increase in the number ($3L - 1$) of parameters accounted for by the Schwarz criterion in table 9.

In sum, the sequential mixture of normal distributions model is the most likely alternative for the distribution of stock returns followed by the Student t, the generalized mixture of normal distributions, the Poisson jump process, the stationary normal, and the quarterly partition.

5. Summary of the results and implications

Previous mixture of independent normal distributions models in the literature that attempt to explain the observed kurtosis and skewness in daily stock return data require computationally intensive search procedures and do not identify observations with the parameters of their respective distributions. The sequential (or time-ordered) mixture of normal distributions model spefication put forth in this paper contains the constraint that there be at least two consecutive observations from the same distribution. With this minimal time-ordered structure, Kim's (1991) Bayesian estimation procedure provides computational simplicity and the direct estimation of change points for parameter nonstationarity. The methodology was applied to daily return data on each of the 30 stocks in the Dow Jones Industrial Average, the Standard & Poors 500 stock index, and the CRSP equal-weighted and CRSP value-weighted indexes. The results indicate that the sequential mixture specification has substantially more descriptive ability than all of the competing alternatives.

Table 9. Asymptotic log posterior odds of each model relative to the i.i.d. normal distribution model with daily returns from July 2, 1962 to December 31, 1990.

Firm's I.D. No.	SM[1]	Student t	Poisson Jump	Two Normals	Three Normals	Four Normals	Five Normals	Quarterly Partition
1	391.32	372.96	343.09	298.46	294.31	295.41	290.89	218.49
2	454.78	382.81	277.56	334.06	373.83	372.63	364.11	−377.11
3	316.34	235.41	207.90	199.41	226.53	N.C.	207.31	−75.61
4	1,213.75	774.41	759.61	701.23	789.62	790.06	N.C.	247.19
5	958.32	759.77	724.96	690.50	758.04	758.87	N.C.	49.96
6	394.06	348.63	N.C.	338.39	357.41	N.C.	337.98	−411.77
7	453.52	510.15	478.40	457.57	489.65	485.67	N.C.	−374.52
8	683.75	443.30	436.80	420.60	436.23	N.C.	N.C.	−155.96
9	835.12	677.09	637.38	603.47	655.45	N.C.	N.C.	−46.77
10	630.55	525.45	491.77	463.65	522.30	N.C.	N.C.	−172.82
11	552.57	394.82	397.83	366.47	399.38	394.15	389.44	−243.91
12	818.41	637.85	589.27	542.81	N.C.	N.C.	518.46	−59.24
13	669.40	520.15	473.31	440.87	N.C.	N.C.	N.C.	−183.56
14	565.64	382.07	365.43	352.91	362.52	357.38	344.93	−226.14
15	649.84	438.61	409.25	381.97	423.88	419.40	406.51	−248.65
16	577.74	485.14	466.23	397.25	408.83	N.C.	459.42	−297.64
17	555.08	392.19	364.88	337.46	N.C.	N.C.	N.C.	−302.65
18	419.56	431.67	404.71	342.04	422.72	N.C.	N.C.	−419.80
19	737.94	469.34	447.42	438.82	450.58	442.96	431.52	17.79
20	707.92	612.65	568.17	574.30	565.18	N.C.	N.C.	95.00
21	236.64	264.35	N.C.	238.79	N.C.	248.05	N.C.	−467.77
22	551.29	493.29	457.22	447.83	480.29	N.C.	N.C.	−278.62
23	344.37	317.44	303.90	N.C.	310.77	304.55	292.32	−436.61
24	867.72	726.29	677.32	669.16	656.75	N.C.	666.37	36.12
25	826.94	619.06	592.09	516.98	596.97	604.59	591.94	−47.50
26	975.21	632.68	593.73	N.C.	620.30	610.34	N.C.	92.63
27	895.12	676.22	642.63	615.46	668.35	668.48	658.30	5.71
28	344.18	303.19	294.43	281.41	295.08	285.99	N.C.	−463.87
29	658.20	702.76	660.29	567.31	692.44	688.99	N.C.	−197.84
30	443.50	449.22	457.76	417.62	461.96	N.C.	443.43	−375.89
S&P	1,223.08	735.82	668.02	663.63	703.44	N.C.	N.C.	364.34
EW	1,443.23	1,009.04	965.13	913.87	1,006.32	1,007.33	994.99	597.26
VW	1,170.49	715.19	646.37	643.60	667.61	656.94	N.C.	323.75

Note: N.C. indicates no convergence.

[1]Sequential mixture of normal distributions.

Note that the sequential mixture of normal distributions model estimates parameter shift dates without a priori knowledge of event dates. This methodology may be particularly useful in security price performance tests. The estimated change points reflect information events. This is important in capturing the effects of multiple announcements, information

leakage, and price revelations from informed traders. These problems arise in a merger or acquisition bidding contest or when measuring the effect of tax or regulatory policy changes. Furthermore, if the information also results in a change in equilibrium expected returns, tests of abnormal performance require the estimates of the true mean and variance for each day surrounding the event to construct appropriate inferences.

The sequential mixture model may also be useful for option valuation. The Black and Scholes (1973) European option pricing model tends to exhibit a systematic bias when employed to value American call options. In particular, Geske and Roll (1984) suggest that the variance bias in Black and Scholes (1972) could be attributed to nonstationary stock volatility. Merton (1976a, 1976b) derives an option pricing model based on a Poisson jump-diffusion process. Although Ball and Torous (1985) confirmed the presence of jumps in security returns, they did not fiond any operationally significant differences between the Black-Scholes and Merton model call option prices. In this paper, we provide evidence that the sequential mixture model is considerably more descriptive of the actual distribution of stock returns than the jump-diffusion process. Therefore, additional research is required to determine whether an estimate of the standard deviation from the current regime of the sequential mixture model results in unbiased Black-Scholes option prices.

Acknowledgments

Appendix

A.1. Derivation of the joint posterior distribution of change points

The likelihood function of the sequential mixture of normal distributions model specification is

$$
L(\Theta \mid R) = (2\pi)^{-T/2} \prod_{i=1}^{K} \left\{ r_i^{(\tau_i - \tau_{i-1})/2} \exp \left\{ -\frac{r_i}{2} \sum_{t=\tau_{i-1}+1}^{\tau_i} (R_t - \mu_i)^2 \right\} \right\},
$$

where $r_i = 1/\sigma_i^2$, $i = 1, \ldots, K$, and $R = (R_1, \ldots, R_T)$ is an observed sotck return vector. Without loss of generality, we assume $\tau_0 = 0$ and $\tau_K = T$.

The prior distribution of $\Theta = (\mu_1, \ldots, \mu_K, \sigma_1^2, \ldots, \sigma_K^2, \tau_1, \ldots, <_{K-1})$ is

$$
\pi(\Theta) \propto \pi(\Psi) \frac{1}{r_1 r_2 \ldots r_K},
$$

where $\Psi = (\tau_1, \ldots, \tau_{K-1})$ and $\pi(\Psi)$ is a joint prior distribution of change points. The notation $\propto$ means proportional to. Then, the posterior distribution of Θ becomes

$$\pi(\Theta \mid R) \propto \pi(\Psi) \prod_{i=1}^{K} \left[r_i^{(\tau_i - \tau_{i-1})/2 - 1} \exp \left\{ -\frac{r_i}{2} \sum_{t=\tau_{i-1}+1}^{\tau_i} (R_t - \mu_i)^2 \right\} \right]. \tag{A1}$$

By using a gamma pdf and a normal pdf, the joint posterior distribution of the change points is obtained by integrating out equation (A1) with respect to $(\mu_1, \ldots, \mu_K, \sigma_1, \ldots, \sigma_K)$ to yield

$$\pi(\Psi \mid R) \propto \pi(\Psi) \prod_{i=1}^{K} \left[\int_{\mu_i} \int_{r_i} r_i^{(\tau_i - \tau_{i-1})/2 - 1} \exp \left\{ -\frac{r_i}{2} \sum_{t=\tau_{i-1}+1}^{\tau_i} (R_t - \mu_i)^2 \right\} dr_i d\mu_i \right]$$

$$\propto \pi(\Psi) \prod_{i=1}^{K} \left[(\tau_i - \tau_{i-1})^{-1/2} \Gamma \left[\frac{\tau_i - \tau_{i-1} - 1}{2} \right] \left\{ \sum_{t=\tau_{i-1}+1}^{\tau_i} (R_t - \bar{R}_i)^2 \right\}^{\tau_i - \tau_{i-1} - 1/2} \right] \tag{A2}$$

By setting $K = 2$, we obtain the posterior distribution of τ of equation (2).

A.2. Derivation of the posterior distribution of $\delta = \mu_1 - \mu_2$ and $\rho = \sigma_2^2/\sigma_1^2$

By using the conditionality, the unconditional posterior distribution of δ is defined as

$$\pi(\delta \mid R) = \sum_{\tau} \left\{ \int \pi(\delta \mid \rho, \tau, R) \pi(\rho \mid \tau, R) d\rho \right\} \pi(\tau \mid R). \tag{A3}$$

Likewise, the unconditional posterior distribution of ρ is also defined as follows;

$$\pi(\rho \mid R) = \sum_{\tau} \pi(\rho \mid \tau, R) \pi(\tau \mid R). \tag{A4}$$

Therefore, we need to obtain the conditional posterior distributions of δ and ρ for their unconditional distributions.

The joint posterior distribution of δ, ρ, and τ is given by

$$\pi(\tau, \delta, \rho \mid R) = \int_{\mu_2} \int_{\rho_2} \pi_2(\tau, \delta + \mu_2, \mu_2, \rho r_2, r_2 \mid R) \mid r_2 \mid dr_2 d\mu_2, \tag{A5}$$

where π_2 is the functional form of equation (A1) when $K = 2$. More precisely, the functional for $\pi_2(\cdot)$ is

$$\pi_2(\tau, \, \delta \, + \, \mu_2, \, \mu_2, \, \rho r_2, \, r_2 | R) \, \propto \, \pi(\tau) \, \rho^{\tau/2-2} r_2^{T/2-1} \, \exp \left\{ - \frac{r_2}{2} \, (\rho \beta_1^* + \beta_2^*) \right\},$$

where

$$\beta_1^* = \sum_{t=1}^{\tau} [R_t - (\delta + \mu_2)]^2, \qquad \beta_2^* = \sum_{t=\tau+1}^{n} (R_t - \mu_2)^2.$$

Now by using an inverted gamma pdf, we simplify the interior integration part of equation (A5) as follows:

$$\int_{r_2} \pi_2(\tau, \, \delta \, + \, \mu_2, \, \mu_2, \, \rho r_2, \, r_2 | R) \, |r_2| \, dr_2 \, = \, \pi(\tau) \rho^{\tau/2-1} \Gamma(n/2)(\rho \beta_1^* + \beta_2^*)^{-n/2}.$$

Then, the joint posterior distribution of τ, δ, and ρ is given by

$$\pi(\tau, \, \delta, \, \rho | R) \, \propto \, \pi(\tau) \rho^{\tau/2-1} \Gamma(n/2) \int_{\mu_2} (\rho \beta_1^* + \beta_2^*)^{-n/2} \, d\mu_2$$

$$\propto \, \pi(\tau) \rho^{n_1/2-1} \Gamma[(n-1)/2](\rho n_1 + n_2)^{-1/2}(\rho S_1 + S_2)^{-(n-1)/2}$$

$$\left\{ 1 + \frac{\rho n_1 n_2}{(\rho S_1 + S_2)(\rho n_1 + n_2)} \, [\delta - (\bar{R}_1 - \bar{R}_2)]^2 \right\}^{-(n-1)/2}, \qquad (A6)$$

where $n_1 = \tau$, $n_2 = n - \tau$, $S_1 = \Sigma_{t=1}^{\tau}(R_t - \bar{R}_1)^2$, and $S_2 = \Sigma_{t=\tau+1}^{n}(R_t - \bar{R}_2)^2$. Note that all constants with respect to τ, δ, and ρ are dropped.

By transforming $\rho' = \rho \{S_1/(n_1 - 1)\}/\{S_2/(n_2 - 1)\} = \rho(\hat{\sigma}_1^2/\hat{\sigma}_2^2)$, the joint posterior distribution of δ, ρ, and τ in equation (A6) becomes decomposed as follows:

$$\pi(\tau, \, \delta, \, \rho | R) \, \propto \, \tau(\tau) \tau^{-1/2}(n - \tau)^{-1/2} \Gamma \left(\frac{\tau - 1}{2} \right) \Gamma \left(\frac{n - \tau - 1}{2} \right) S_1^{-(\tau-1)/2} S_2^{-(n-\tau-1)/2}$$

$$\frac{\Gamma((n-2)/2)}{\Gamma((\tau - 1)/2)\Gamma((n - \tau - 1)/2)} \left(\frac{\tau - 1}{n - \tau - 1} \right)^{(\tau-1)/2} (\rho')^{(\tau-1)/2-1}$$

$$\left(1 + \frac{\tau - 1}{n - \tau - 1} \rho' \right)^{-(n-2)/2} \left(1 + \frac{\rho\tau(n - \tau)[\delta - (\bar{R}_1 - \bar{R}_2)]^2}{(\rho S_1 + S_2)(\rho\tau + n - \tau)} \right)^{-(n-1)/2}$$

$$\equiv \, \pi(\tau | R) \, \pi(\rho' | \tau, R) \, \pi(\delta | \rho, \rho, R).$$

Therefore, the conditional posterior distribution of δ given ρ and τ, $\pi(\delta | \tau, \rho, R)$, is the Student t distribution with location parameter $\hat{\delta} = \bar{R}_1 - \bar{R}_2$, scale parameter $(\rho S_1 + S_2)$ $(\rho\tau + n - \tau)/[(n - 2)\rho\tau(n - \tau)]$, and $(n - 2)$ debgrees of freedom, and the conditional posterior distribution of ρ given τ, $\pi(\rho' | \tau, R)$, is an F distribution with $(\tau - 1, n - \tau - 1)$ degrees of freedom. The posterior distribution of τ is defined in equation (2).

A.3. HPD interval test for $\delta = 0$ and $\rho = 1$

The conditional p-value for H_{01} given ρ and τ is defined as

$$p_{\delta=0|\rho,\tau} = E_{\bar{C}_\delta}[\pi(\delta|\rho, \tau, R)],$$

where $E_{\bar{C}_\delta}[\pi(\cdot)]$ indicates the area of a posterior pdf $\pi(\cdot)$ over the region $\bar{C}_\delta$, and $\bar{C}_\delta$ is the compliment of C_δ. C_δ is the $(1 - \alpha)$ credible set (i.e., the HPD interval) of δ defined as

$$C_\delta = \{\delta : \pi_{\alpha/2}(\delta|\rho, \tau, R) \leq \delta \leq \pi_{1-\alpha/2}(\delta|\rho, \tau, R)\},$$

where $\pi_{\alpha/2}(\delta|\cdot)$ is the $\alpha/2$th quantile of the posterior distribution of δ, $\pi(\delta|\cdot)$. Therefore, the unconditional p-value for H_{01} is defined as

$$p_{\delta=0} = \sum_\tau \left\{ \int_\rho E_{\bar{C}_\delta}[\pi(\delta|\rho, \tau, R)]\pi(\rho|\tau, R)d\rho \right\} \pi(\tau|R)$$

$$= \sum_\tau \left\{ \int 2(1 - \Im_{n-2}(|t(0)|))\pi(\rho|\tau, R) \right\} \pi(\tau|R), \tag{A7}$$

where

$$t(\delta) = \frac{\delta - (\bar{R}_1 - \bar{R}_2)}{(\rho S_1 + S_2)(\rho\tau + n - \tau)/[(n - 2)\rho\tau(n - \tau)]}.$$

In fact, when $\delta = 0$ and $\rho = 1$, the statistic $t(0)$ is the standard t-test statistic with $(n - 2)$ degrees of freedom for the null hypothesis that two means are equal assuming the same variance. Basically, $2(1 - \Im_{n-2}(|t((0)|))$ in equation (A7) is the p-value for the null H_{01} by the t-test given the ratio of two variances, ρ, and τ. That is, it equals $p_{\delta=0|\rho,\tau}$. Based on Box and Tiao (1973), the inside bracket of equation (A7) can be approximated as the standard Student t-test and we obtain the approximate unconditional p-value of equation (1).

Likewise, the unconditional p-value for the null H_{02} is defined as follows:

$$p_{\rho=1} = \sum_\tau E_{\bar{C}_\rho}[\pi(\rho|\tau, R)]\pi(\tau|R)$$

$$= \sum_\tau 2\{1 - \Upsilon_{\tau-1,n-\tau-1}(F(1))\} \pi(\tau|R).$$

Notes

1. Additional evidence inconsistent with the stable Paretian hypothesis can be found in Hagerman (1978), Akgiray and Booth (1988), Lau et al. (1990), and Tucker (1992).

2. For some evidence, see Officer (1972), Boness et al. (1974), Hsu et al. (1974), Beaver (1968), Patell and Wolfson (1981), Christie (1982), Conrad and Kaul (1988) and Bodurtha and Mark (1991).

3. The weakness of a day-of-the-weak effect in daily data is also supported by Connolly (1989) and Akgiray (1989).

4. A posteriori means after classification of observations into each regime, while a prior means before classification of observations. The generalized normal mixture model has one probability density function. The classification of observations into each population is not part of the estimation procedure. If both the generalized discrete mixture and its classification praoblem are considered together, this posterior version would contain the sequential mixture of normal distributions model as a subset.

5. See DeGroot (1970) for further explanation about the generalized maximum likelihood estimator.

6. The HPD interval is defined as follows: Let $\pi(\theta|\text{data})$ be a posterior density function of θ. A region C in the parameter space of θ is called an HPD interval of content $(1 - \alpha)$ if $Pr\{\theta \in C|\text{data}\} = 1 - \alpha$ and, for $\theta_1 \in C$ and $\theta_2 \notin C$, $\pi(\theta_1|\text{data}) \geq \pi(\theta_2|\text{data})$. The HPD interval test is different from the traditional Bayes testing procedure carried out within a posterior odds framework. The typical Bayes test presumes that the null value of a target parameter is believed to be stronger than any other value near the null value, a presumption which is sometimes true, but in many applications unsupported. Lindley (1965) suggests a Bayesian significance test by the HPD interval and emphasizes that this type of significance test is appropriate only for circumstances in which prior knowledge of the target parameter is vague or diffuse, and one of the hypotheses is a single point. This method is similar to the sampling theory approach of rejecting a null hypothesis when the hypothesized value for a parameter falls outside a confidence interval. The confidence interval corresponds to the HPD interval in the HPD interval test. If the posterior distribution of a target parameter is concentrated on the hypothesized value of the parameter, then it is hard to reject the null hypothesis.

7. Although the HPD interval test is similar to the sampling theoretic approach, the concept and results are different. Kim (1991) shows that the HPD interval test for stationarity of regression parameters outperforms in power the conventional non-Bayesian techniques such as the cusum and cusum of squares tests, even if the diffuse prior is used.

8. With our assumption of conditional (on t) multivariate normality, portfolios are stable under addition. That is, linear combinations of normally distributed stock returns are also normally distributed.

References

Akgiray, V., "Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts." *Journal of Business* 62, pp. 55–80, (1989).

Akgiray, V. and G.G. Booth, "Stock Price Processes with Discontinuous Time Paths: An Empirical Examination." *Financial Review* 21, pp. 163–184, (1984).

Akgiray, V. and G.G. Booth, "Compound Distribution Models of Stock Returns: An Empirical Comparison." *Journal of Financial Research* 10, pp. 269–279, (1987).

Akgiray, V. and G.G. Booth, "The Stable-Law Model of Stock Returns." *Journal of Business and Economic Statistics* 6, pp. 51–57, (1988).

Ball, C.A. and W.N. Torous, "On Jumps in Common Stock Prices and Their Impact on Call Option Pricing." *Journal of Finance* 40, pp. 155–173, (1985).

Ball, C.A. and W.N. Torous, "Investigating Security-Price Performance in the Presence of Event-Date Uncertainty." *Journal of Financial Economics* 22, pp. 123–153, (1988).

Beaver, W., "The Information Content of Annual Earnings Announcements." *Empirical Research in Accounting: Selected Studies, Supplement to Journal of Accounting Research* 6, pp. 67–92, (1968).

Black, F. and M. Scholes, "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81, pp. 637–659, (1973).

Black, F. and M. Scholes, "The Valuation of Option Contracts and a Test of Market Efficiency." *Journal of Finance* 27, pp. 399–417, (1972).

Blattberg, R.C. and N.J. Gonedes, "A Comparison of the Stable and Student Distribution as Statistical Models for Stock Prices." *Journal of Business* 47, pp. 244–280, (1974).

Bodurtha, J.N. and N.C. Mark, "Testing the CAPM with Time-Varying Risks and Returns." *Journal of Finance* 46, pp. 1485–1505, (1991).

Boness, A. J., A. Chen, and S. Jatusipitak, "Investigations of Nonstationarity in Stock Prices." *Journal of Business* 47, pp. 518–537, (1974).

Box, G.E.P. and G.C. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.

Christie, A., "The Stochastic Behavior of Common Stock Variances: Value, Leverage, and Interest Rate Effects." *Journal of Financial Economics* 10, pp. 407–432, (1982).

Christie, A. "On Information Arrival and Hypothesis Testing in Event Studies." Working Paper, University of Rochester.

Connolly, R.A. "An Examination of the Robustness of the Weekend Effect." *Journal of Financial and Quantitative Analysis* 24, pp. 133–169, (1989).

Conrad, J. and G. Kaul, "Time-Variation in Expected Returns." *Journal of Business* 61, pp. 409–425, (1988).

DeGroot, M.H., *Optimal Statistical Decisions*. New York: McGraw-Hill Book Co., 1970.

Fama, E.F., "The Behavior of Stock Market Prices." *Journal of Business* 38, pp. 34–105, (1965).

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

Geske, R. and R. Roll, "On Valuing American Call Options with the Black-Scholes European Formula." *Journal of Finance* 39, pp. 443–455, (1984).

Hagerman, R., "More Evidence on the Distribution of Security Returns." *Journal of Finance* 33, pp. 1213–1220, (1978).

Hsu, D.A., R. Miller, and D. Wichern, "On the Stable Paretian Behavior of Stock Market Prices." *Journal of American Statistical Association* 69, pp. 108–113, (1974).

Jarrow, R. and E. Rosenfeld, "Jump Risk and the Intertemporal Capital Asset Pricing Model." *Journal of Business* 57, pp. 337–351, (1984).

Kim, D., "A Bayesian Significance Test on the Stationarity of Regression Parameter." *Biometrika* 78, pp. 667–675, (1991).

Kon, S.J., "Models of Stock Returns—A Comparison." *Journal of Finance* 39, pp. 147–165, (1984).

Lau, A., H. Lau, and J. Wingender, "The Distribution of Stock Returns: New Evidence Against the Stable Model." *Journal of Business and Economic Statistics* 8, pp. 217–223, (1990).

Lindley, D.V., *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference*, Cambridge: Cambridge University Press, 1965.

Mandelbrot, B., "New Method in Statistical Economics." *Journal of Political Economy* 71, pp. 421–440, (1963a).

Mandelbrot, B., "The Variation of Cedrtain Speculative Prices." *Journal of Business* 36, pp. 394–419, (1963b).

Mandelbrot, B., "The Variation of Some Other Speculative Prices." *Journal of Business* 40, pp. 393–413, (1967).

Merton, R.C., "Option Pricing When Underlaying Stock Returns are Discontinuous." *Journal of Financial Economics* 3, pp. 125–144, (1976a).

Merton, R.C., "The Impact on Option Pricing of Specification Error in the Underlying Stock Price Returns." *Journal of Finance* 31, pp. 333–350, (1976b).

Officer, R., "The Distribution of Stock Returns." *Journal of American Statistical Association* 67, pp. 807–812, (1972).

Patell, J. and M. Wolfson, "The Ex Ante and Ex Post Price Effects of Quarterly Earnings Announcements Reflected in Option and Stock Prices." *Journal of Accounting Research* 19, pp. 434–458, (1981).

Press, S.J., "A Compound Events Model for Security Prices." *Journnal of Business* 45, pp. 317–335, (1968).

Schwarz, G., "Estimating the Dimension of a Model." *Annals of Statistics* 6, pp. 461–464, (1978).

Tucker, A.L., "A Reexamination of Finite and Infinite-Variance Distributions as Models of Daily Stock Returns." *Journal of Business and Economic Statistics* 10, pp. 73–81, (1992).