Use of Cumulative Sums of Squares for Retrospective Detection of Changes of Variance

Carla INCLÁN and George C. TIAO*

This article studies the problem of multiple change points in the variance of a sequence of independent observations. We propose a procedure to detect variance changes based on an iterated cumulative sums of squares (ICSS) algorithm. We study the properties of the centered cumulative sum of squares function and give an intuitive basis for the ICSS algorithm. For series of moderate size (i.e., 200 observations and beyond), the ICSS algorithm offers results comparable to those obtained by a Bayesian approach or by likelihood ratio tests, without the heavy computational burden required by these approaches. Simulation results comparing the ICSS algorithm to other approaches are presented.

KEY WORDS: Cumulative sum of squares; Multiple change points; Variance change.

1. INTRODUCTION

We study the detection of multiple changes of variance in a sequence of independent observations. There are series, particularly in the area of finance, that do not follow the usual assumption of constant variance underlying most models for time series. We consider series that present a stationary behavior for some time, then suddenly the variability of the error term changes; it stays constant again for some time at this new value, until another change occurs.

The statistical literature on changes of variance started with Hsu, Miller, and Wichern (1974), who offered this formulation as an alternative to the Pareto distribution to model stock returns. Previous studies had pointed out the nonnormality of these series and suggested the use of a heavy-tailed distribution. Hsu, Miller, and Wichern proposed a normal probability model with a nonstationary variance subject to step changes at irregular time points.

There are many works aimed at identifying the point of change in a sequence of independent random variables (Hinkley 1971; Menzefricke 1981; Smith 1975, 1980). Booth and Smith (1982) used the Bayes ratio to decide whether a series presents a single change of variance at an unknown point. Hsu (1977, 1979, 1982) studied the detection of a variance shift at an unknown point in a sequence of independent observations, focusing on the detection of points of change one at a time because of the heavy computational burden involved in looking for several points of change simultaneously. Worsley (1986) used maximum likelihood methods to test a change in mean for a sequence of independent exponential family random variables, to estimate the change point, and to give confidence regions. His work focused on finding one change point at a time.

For autocorrelated observations, Wichern, Miller, and Hsu (1976) studied an autoregressive model of order one, having a sudden variance change at an unknown point. Abraham and Wei (1984) used a Bayesian framework to solve the same problem. Baufays and Rasson (1985) estimated the variances and the points of change of maximum likelihood. Their method handles several points of change simultaneously, and improves on Wichern et al. by reducing the computational effort involved. Tsay (1988) discussed autoregressive moving average models allowing for outliers and variance changes and proposed a scheme for finding the point of variance change. Broemeling and Tsurumi (1987) studied structural change in econometric models using Bayesian techniques. Applications of change point models in areas other than finance include the work of Cobb (1978), Commenges, Seal, and Pinatel (1986), Haccou, Meelis, and van de Geer (1988), and Haccou and Meelis (1988).

The approach presented here uses cumulative sums of squares to search for change points systematically at different pieces of the series. It is based on a centered version of the cumulative sum of squares presented by Brown, Durbin, and Evans (1975). The search is done following an algorithm to find multiple change points in an iterative way.

This article is organized as follows. Section 2 presents the centered cumulative sum of squares function D_k, its relationship to the F statistic used to test equality of variances, the expected value of D_k under changes of variance, and the asymptotic distribution of $\max_k |D_k|$ when the series has homogeneous variance. Section 3 introduces the ICSS algorithm and illustrates it using a financial time series. Section 4 reports a simulation study comparing the performance of the ICSS algorithm to other approaches and illustrates the cost of using the ICSS algorithm in comparison to other approaches, measuring cost in terms of the CPU time.

2. CENTERED CUMULATIVE SUMS OF SQUARES

The main interest is to study the variance of a given sequence of observations retrospectively, so we can use all the information on the series to indicate the points of variance change.

Let $C_k = \sum_{t=1}^{k} a_t^2$ be the cumulative sum of squares of a series of uncorrelated random variables $\{a_t\}$ with mean 0 and variances σ_t^2, $t = 1, 2, \ldots, T$. Let

$$D_k = \frac{C_k}{C_T} - \frac{k}{T}, \quad k = 1, \ldots, T, \quad \text{with} \quad D_0 = D_T = 0 \qquad (1)$$

be the centered (and normalized) cumulative sum of squares. The plot of D_k against k will oscillate around 0 for series with homogeneous variance. When there is a sudden change

* Carla Inclán is Assistant Professor, School of Business Administration, Georgetown University, Washington DC 20057. George C. Tiao is W. Allen Wallis Professor of Statistics, Graduate School of Business, University of Chicago, IL 60637.

Journal of the American Statistical Association
September 1994, Vol. 89, No. 427, Theory and Methods

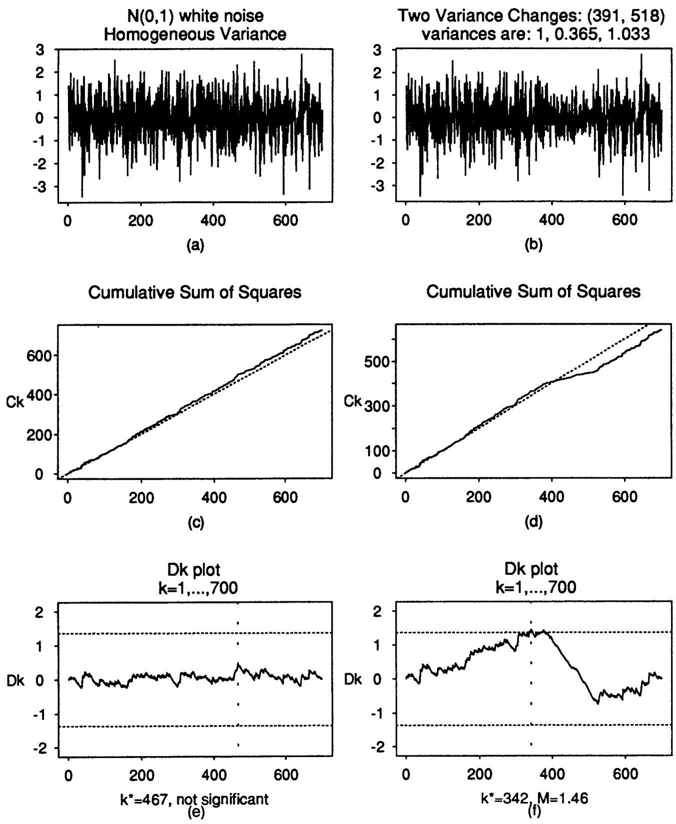

Figure 1. Examples of Cumulative Sums of Squares Plots.

in variance, the plot of D_k will exhibit a pattern going out of some specified boundaries with high probability. These boundaries can be obtained from the asymptotic distribution of D_k assuming constant variance.

Figure 1a shows a series of $N(0, 1)$ white noise with constant variance σ^2; Figure 1b presents the same series with two changes of variance, at $t = 391$ and $t = 518$. The variances are $\sigma_t^2 = \tau_0^2 = 1$, $t = 1, \ldots, 390$; $\sigma_t^2 = \tau_1^2 = .365$, $t = 391, \ldots, 517$; and $\sigma_t^2 = \tau_2^2 = 1.033$, $t = 518, \ldots, 700$. Figure 1, c and d shows the C_k function, and Figure 1, e and f illustrates the D_k function for the two series presented in a and b. For the series with homogeneous variance, C_k presents roughly a straight line with slope $\sigma^2 = 1$. When there are changes in the variance, the plot appears as a broken line consisting of several straight pieces.

The plot of D_k presents a better picture, because when there is a change in variance the slope of C_k changes slightly, being positive all the time, whereas the slope of D_k shows a drastic change, even a change of sign, creating a peak or a trough according to whether the variance changes to a smaller or a greater value. Besides, the horizontal is an easier point of visual reference than a straight line with positive slope. This behavior leads to the search for a variance change point via $\max_k |D_k|$. Let k^* be the value of k at which $\max_k |D_k|$ is attained. If this maximum absolute value exceeds a predetermined boundary, then we may conclude that there is a change point near k^* and take k^* as an estimate of the change point.

Under variance homogeneity, $\sqrt{T/2}D_k$ behaves like a Brownian bridge asymptotically (see Sect. 2.3). From Table 1, the asymptotic critical value is $D^*_{.05} = 1.358$. The two horizontal dashed lines in Figure 1, e and f are at ± 1.358, and we can detect in Figure 1f that the maximum of $\sqrt{T/2}|D_k|$ exceeds the boundary value at $k^* = 342$.

2.1 Relationship of D_k to the F Statistic and the Likelihood Ratio

For a fixed k, the value of D_k can be written as a function of the usual F statistic for testing equality of variances between two independent samples. Specifically, let the first sample consist of observations a_i, $i = 1, \ldots, k$, with variance τ_0^2, and let the second sample be a_j, $j = k + 1, \ldots, T$, with variance τ_1^2. Then the F statistic for testing H_0: $\tau_0^2 = \tau_1^2$ against H_a: $\tau_0^2 < \tau_1^2$ is $F_{T-k,k} = ((C_T - C_k)/(T - k))/(C_k/k)$.

Thus D_k can be expressed in terms of $F_{T-k,k}$ as

$$D_k = \frac{C_k}{C_T} - \frac{k}{T} = \frac{k(T-k)}{T^2}\left[\frac{1 - F_{T-k,k}}{\dfrac{k}{T} + \dfrac{T-k}{T}F_{T-k,k}}\right]. \quad (2)$$

For a fixed k, $D_k(F)$ is a monotone function of F; it depends only on k through k/T. Note the important distinction: The F statistic is used with known k, whereas we will be looking for $\max_k |D_k|$ to determine the location of the change point.

If we assume that the $\{a_t\}$ are Normally distributed, with mean 0 and variances σ_t^2, $t = 1, \ldots, T$, then we can obtain the likelihood ratio for testing one change against no change in the variance. Let $N_T = 0$ represent the case of no change and let $N_T = 1$ represent the case of one change. The concentrated log-likelihood for $N_T = 0$ is

$$l(N_T = 0; \mathbf{a}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log(T^{-1}C_T) - \frac{T}{2}, \quad (3)$$

where $\mathbf{a} = (a_1, \ldots, a_T)'$. Now let κ be the point of change. Then the concentrated log-likelihood function for $N_T = 1$ and κ is

$$l(\kappa, N_T = 1; \mathbf{a}) = -\frac{T}{2}\log(2\pi) - \frac{\kappa}{2}\log\left(\frac{1}{\kappa}C_\kappa\right) - \frac{T-\kappa}{2}\log\left(\frac{1}{T-\kappa}(C_T - C_\kappa)\right) - \frac{T}{2}. \quad (4)$$

Table 1. Empirical and Asymptotic Quantiles of $\max_k \sqrt{T/2}|D_k|$

T	100		200		300		400		500		∞
p	q_D	SE	q_D	SE	q_D	SE	q_D	SE	q_D	SE	D^*_{1-p}
.05	.44	.003	.47	.003	.47	.003	.48	.003	.049	.003	.520
.10	.50	.003	.52	.003	.53	.003	.53	.003	.054	.002	.571
.25	.60	.004	.63	.003	.63	.003	.64	.003	.065	.003	.677
.50	.75	.004	.78	.003	.78	.003	.79	.003	.080	.003	.828
.75	.94	.004	.97	.004	.97	.004	.97	.004	1.00	.004	1.019
.90	1.14	.006	1.16	.006	1.18	.007	1.18	.006	1.20	.006	1.224
.95	1.27	.009	1.30	.004	1.31	.008	1.31	.010	1.33	.009	1.358
.99	1.52	.004	1.55	.012	1.57	.028	1.57	.020	1.60	.018	1.628

NOTE: Estimated from 10,000 replicates of series of T independent $N(0, 1)$ observations. D^*_{1-p} is defined by $P\{\sup_r |W^0_r| < D^*_{1-p}\} = p$.

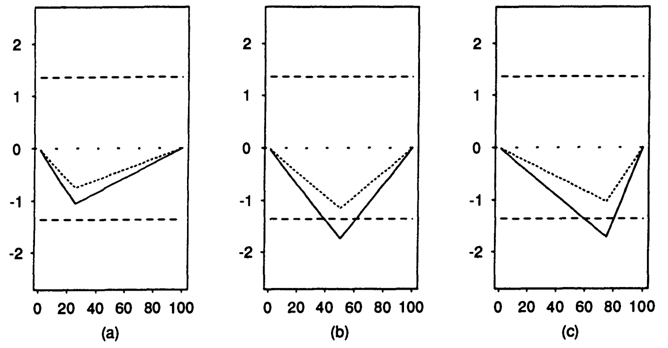

Figure 2. Approximate Expected Value of $\sqrt{T/2}D_k$ for Series With One Change Point. Change points are (a) $\kappa = .25T$, (b) $\kappa = .50T$ (c) $\kappa = .75T$. Two values of Δ are shown on each plot: dotted line, $\Delta = 2$; solid line, $\Delta = 3$.

The maximum likelihood estimator of κ is obtained by maximizing this function with respect to κ. Hence the logarithm of the likelihood ratio for testing $N_T = 0$ against $N_T = 1$ is

$$LR_{0,1} = -\max_{\kappa}\left\{-\frac{\kappa}{2}\log\left(1 + \frac{T}{\kappa}D_\kappa\right) - \frac{T-\kappa}{2}\log\left(1 - \frac{T}{T-\kappa}D_\kappa\right)\right\}. \quad (5)$$

Note that although $LR_{0,1}$ and $\max_\kappa|D_\kappa|$ are related, they are not equivalent for finite T. Using a second-order Taylor's expansion of (5) as a function of D_κ, for large κ and $T - \kappa$, $LR_{0,1} \approx -\max_\kappa\{T^3D_\kappa^2/(4\kappa(T - \kappa))\}$. The function $\max_\kappa|D_\kappa|$ puts more weight near the middle of the series, thus inducing k^* to be biased toward $T/2$.

2.2 Expected Value of D_k Given the Variance Changes

In this section we study the behavior of $E[D_k]$, the expected value of D_k, as a function of k. We consider the expected value under different assumptions about the variance of the series. See Appendix A for the derivations.

Consider the case where the series has one change point at κ_1; that is, $\sigma_t^2 = \tau_0^2$ for $t = 1, \ldots, \kappa_1$ and $\sigma_t^2 = \tau_1^2$ for $t = \kappa_1 + 1, \ldots, T$. The value of $E[D_k]$ is, to order $o(T^{-1})$, a piecewise linear function with a change of slope at κ_1. Figure 2 illustrates the influence of the location of the changepoint κ_1 and the value of the ratio of variances $\Delta_1 = \tau_1^2/\tau_0^2$. Figure 2 shows three plots corresponding to the change point at $\kappa_1 = 25$, $\kappa_1 = 50$, and $\kappa_1 = 75$, with two values of Δ_1 for a series of length $T = 100$. The values of $E[D_k]$ have been multiplied by $\sqrt{T/2}$ for comparison with the asymptotic boundaries ± 1.358 also shown in the plots.

Note that the approximate $E[D_k]$ has a symmetry involving both the location of the changepoint and the variance ratio. Let $a[1:T] = \{a_1, a_2, \ldots, a_T\}$ be the original series to which D_k is applied, and let $a[T:1] = \{a_T, a_{T-1}, \ldots, a_1\}$ be the reversed series. Call $g_k(\kappa_1, \Delta_1)$ the approximate expected value of $D_k(a[1:T])$; then the approximate expected value of $D_k(a[T:1])$ is $-g_{T-k}(T - \kappa_1, 1/\Delta_1)$. The implications for the search of the changepoint are important: If the smaller variance corresponds to the shorter segment of the series, then it will be harder to find the changepoint using D_k. This interaction is also present in the F test for difference in variances. The power of the F test is higher for the case of the larger variance with the smaller degrees of freedom. The interaction is less important as the sample size increases.

If there are two changepoints, located at κ_1 and κ_2, with the corresponding variances of τ_0^2 for $t = 1, \ldots, \kappa_1$, τ_1^2 for $t = \kappa_1 + 1, \ldots, \kappa_2$, and τ_2^2 for $t = \kappa_2 + 1, \ldots, T$, then the approximate $E[D_k]$ is a piecewise linear function of k, with changes of slope at κ_1 and κ_2. Figure 3 shows the second-order approximation to $E[D_k]$ when there are two changepoints at $\kappa = (58, 80)$ with $T = 100$. In the figure each of the six plots represents one of the six possible configurations of variances. One configuration has τ_0 as the smallest, τ_1 the intermediate, and τ_2 the largest. In all, there are six ways of permuting the place of the small (S), medium (M), and large (L) variances. Figure 3 clearly illustrates the masking effect when there is more than one changepoint. In most cases these plots show a well-defined peak at one of the two changepoints; that is, the point most likely to be found when the D_k function is applied to a series with two changes of variance. The iterative algorithm presented in Section 3 is designed to lessen the masking effect, as the search for changepoints takes place one by one at different pieces of the series.

2.3 Asymptotic Behavior of D_k Under Homogeneous Variance

The asymptotic distribution of D_k when the random variables $\{a_t\}$ are identically distributed is that of a Brownian

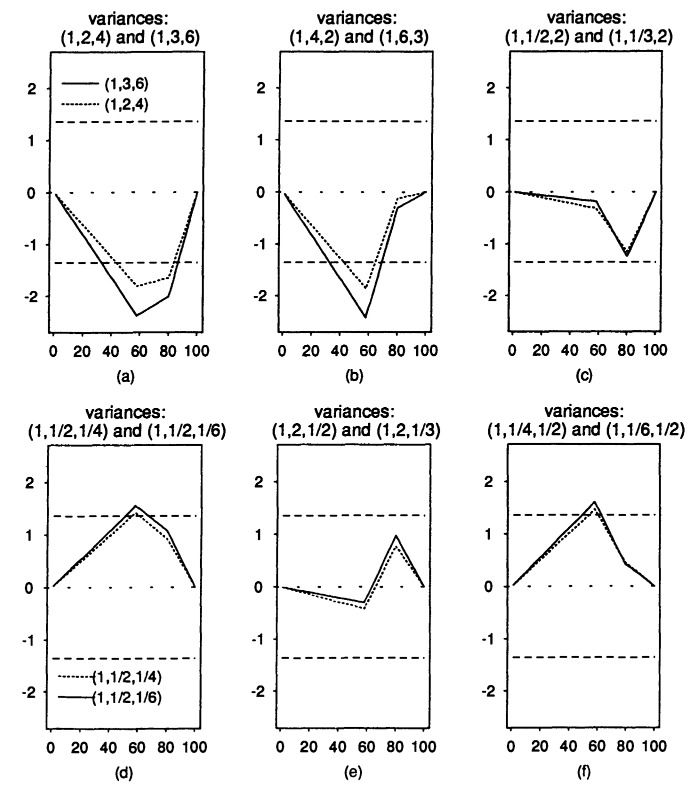

Figure 3. Approximate Expected Value of $\sqrt{T/2}D_k$ for Series with Two Change Points at $(\kappa_1, \kappa_2) = (58, 80)$. Variance ratios are indicated on each plot.

bridge. The notation in this section follows closely that of Billingsley (1968), using n instead of T as the number of observations in the series $\{a_i\}$, $i = 1, \ldots, n$.

Let W represent a Brownian motion process, $E[W_t] = 0$, $E[W_t W_s] = s$, $0 \le s < t$. Let W^0 denote a Brownian bridge, $W_t^0 = W_t - tW_1$, $E[W_t^0] = 0$, $E[W_t^0 W_s^0] = s(1-t)$, $0 \le s < t$, and $W_0^0 = W_1^0 = 0$, with probability 1.

Theorem 1. Let $a_1, a_2, \ldots$ be a sequence of independent, identically distributed Normal$(0, \sigma_a^2)$ random variables. Let $D_k = C_k/C_n - k/n$, where $C_k = \sum_{i=1}^{k} a_i^2$. Then $\sqrt{n/2}D_k \xrightarrow{\mathcal{D}} W^0$. The proof is given in Appendix B.

This asymptotic distribution determines the probability $P\{\sup_t |W_t^0| < D^*\}$, so that asymptotic quantiles of $\max_\kappa |D_\kappa|$ can be obtained. Table 1 compares selected asymptotic quantiles of $\max_\kappa |D_\kappa|$ with corresponding empirical quantiles for various series lengths obtained by simulation.

The standard errors were obtained as $\sqrt{p(1-p)/nf^2}$, where $n = 10{,}000$ is the number of replicates and f is the density at the pth quantile, estimated with the relative frequency per unit for an interval of width .02 centered at the pth quantile. Table 2 gives the percentage of series for which $\max_k \sqrt{T/2}|D_k| < 1.358$. The standard errors in this table are obtained as SE $= \sqrt{\hat{p}(1-\hat{p})/10{,}000}$, where $\hat{p}$ is the proportion of series with $\max_k \sqrt{T/2}|D_k| < 1.358$.

The 95th quantile of $\max_k \sqrt{T/2}|D_k|$ is always lower than the asymptotic value 1.358, thus leading to a smaller empirical type I error when testing for one change of variance. But, for $T \ge 200$, the asymptotic approximation seems to perform well for practical use.

Table 2. Percentage of Series Where $\max_k \sqrt{T/2}|D_k| < 1.358$

T	100	200	300	400	500
% ($\hat{p}$)	97.13	96.51	96.31	96.07	95.53
SE	.167	.184	.189	.194	.207

NOTE: 10,000 replicates, series of T independent $N(0, 1)$ observations. SE $= \sqrt{\hat{p}(1-\hat{p})/10{,}000}$.

3. DETECTIONS OF MULTIPLE CHANGES: THE ITERATED CUMULATIVE SUMS OF SQUARES ALGORITHM

If we were concerned only with the possible existence of a single point of change, then the D_k function would provide a satisfactory procedure. But when we are interested in finding multiple points of variance change on an observed series, the usefulness of the D_k function becomes questionable because of the masking effect. A solution is an iterative scheme based on successive application of D_k to pieces of the series, dividing consecutively after a possible changepoint is found. We now propose a systematic procedure to look for changepoints in an effort to isolate each point. We use the notation $a[t_1 : t_2]$ to represent the series $a_{t_1}, a_{t_1+1}, \ldots, a_{t_2}$, $t_1 < t_2$ and use the notation $D_k(a[t_1 : t_2])$ to indicate the range over which the cumulative sums are obtained.

Iterated Cumulative Sums of Squares (ICSS) Algorithm

Step 0. Let $t_1 = 1$.

Step 1. Calculate $D_k(a[t_1 : T])$. Let $k^*(a[t_1 : T])$ be the point at which $\max_k |D_k(a[t_1 : T])|$ is obtained, and let

$$M(t_1 : T) = \max_{t_1 \le k \le T} \sqrt{(T - t_1 + 1)/2}\;|D_k(a[t_1 : T])|.$$

If $M(t_1 : T) > D^*$, consider that there is a changepoint at $k^*(a[t_1 : T])$ and proceed to Step 2a. The value of D^* is D^*_{1-p} from Table 1 for the desired value of p, usually $p = .95$. If $M(t_1 : T) < D^*$, there is no evidence of variance changes in the series. The algorithm stops.

Step 2a. Let $t_2 = k^*(a[t_1 : T])$. Evaluate $D_k(a[t_1 : t_2])$; that is, the centered cumulative sum of squares applied only to the beginning of the series up to t_2. If $M(t_1 : t_2) > D^*$, then we have a new point of change and should repeat Step 2a until $M(t_1 : t_2) < D^*$. When this occurs, we can say that there is no evidence of variance change in $t = t_1, \ldots, t_2$ and, therefore, the first point of change is $k_{\text{first}} = t_2$.

Step 2b. Now do a similar search starting from the first changepoint found in Step 1, toward the end of the series. Define a new value for t_1: let $t_1 = k^*(a[t_1 : T]) + 1$. Evaluate $D_k(a[t_1 : T])$, and repeat Step 2b until $M(t_1 : T) < D^*$. Let $k_{\text{last}} = t_1 - 1$.

Step 2c. If $k_{\text{first}} = k_{\text{last}}$, there is just one changepoint. The algorithm stops there. If $k_{\text{first}} < k_{\text{last}}$, keep both values as possible changepoints and repeat Step 1 and Step 2 on the middle part of the series; that is, $t_1 = k_{\text{first}} + 1$ and $T = k_{\text{last}}$. Each time that Steps 2a and 2b are repeated, the result can be one or two more points. Call $\hat{N}_T$ the number of changepoints found so far.

Step 3. If there are two or more possible changepoints, make sure they are in increasing order. Let cp be the vector of all the possible changepoints found so far. Define the two extreme values $cp_0 = 0$ and $cp_{\hat{N}_T+1} = T$. Check each possible changepoint by calculating $D_k(a[cp_{j-1} + 1 : cp_{j+1}])$, $j = 1, \ldots, \hat{N}_T$. If $M(cp_{j-1} + 1 : cp_{j+1}) > D^*$, then keep the point; otherwise, eliminate it. Repeat Step 3 until the number of changepoints does not change and the points found in each new pass are "close" to those on the previous pass. In our implementation of this algorithm, we consider that if each changepoint is within two observations of where it was on the previous iteration, then the algorithm has converged. This convergence is achieved in few iterations of Step 3. Note that during each iteration, the newly found points must be kept apart to make an entire pass through the series based on a single set of points.

The actual implementation of this algorithm requires some controls over the number of iterations as a precaution to avoid cycling indefinitely. In the examples analyzed, as well as in extensive simulations performed, the limit of 20 iterations was never attained.

We now illustrate the use of the ICSS algorithm with the series having two variance changes in Figure 1b. From the D_k plot in Figure 1f, we find a possible changepoint at $k^* = 342$. Now, cut the series and consider the first part $a[1 : 342]$. The result is shown in Figure 4a, where the D_k $(a[1 : 342])$ path lies within the boundaries, so that $k_{\text{first}} = 342$. Consider $a[343 : 700]$. The resulting $D_k(a[343 : 700])$ is shown in Figure 4b, where the maximum absolute value

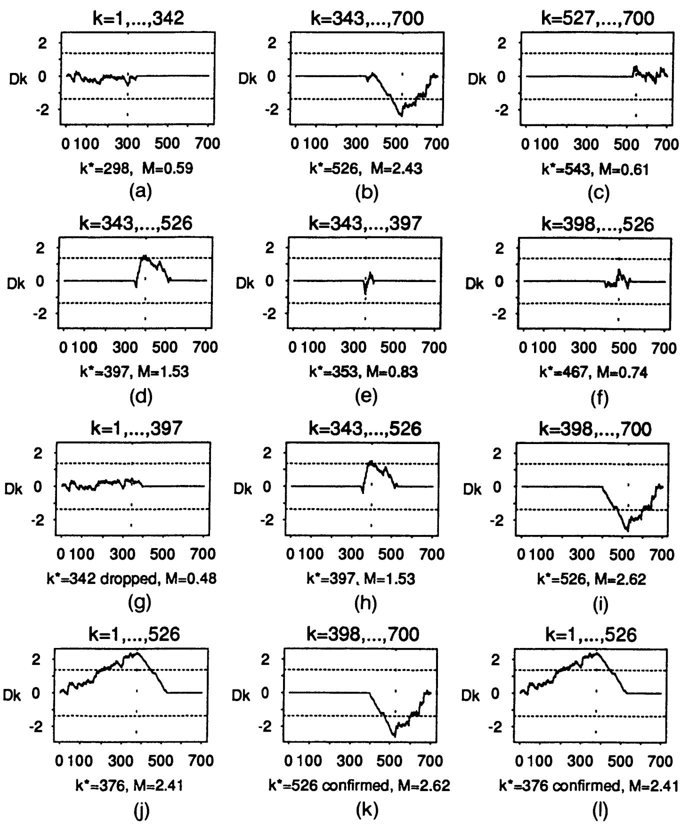

Figure 4. Sequence of D_k Functions From the ICSS Algorithm, Applied to a Series of Independent Observations With Two Changes of Variance.

exceeds the boundary value at $k^* = 526$. Next, the plot in Figure 4c shows no further points of change in the range $a[527 : 700]$ and, therefore, $k_{\text{last}} = 526$. We now repeat the same process for the inner part of the series $a[343 : 526]$ and look for the "first" point of change within these values. Figure 4, d–f illustrates this process. We find one more possible point of change at 397. Plots e and f lie within boundaries. At this point we have three possible points of change: $cp_1 = 342$, $cp_2 = 397$, and $cp_3 = 526$.

This procedure performs a systematic search for possible points of change working from the extremes towards the middle of the series. But it can find too many points. Step 3 can help "fine tune" the location of the points by checking each point given the adjacent ones; that is, checking cp_1 by $D_k(a[1 : 397])$, checking cp_2 by $D_k(a[343 : 526])$, and finally checking cp_3 by $D_k(a[398 : 700])$. This step is repeated until convergence—until the number of points does not change and their locations do not move by more than a specified amount. In our case we use two observations: once each cp_i $i = 1, \ldots, N_T$ is within two units of where it was at the previous iteration, we consider that the algorithm has converged. Figure 4g shows that the point 342 is no longer significant after considering the adjacent ones. We now have $cp_1 = 397$ and $cp_2 = 526$. A second iteration is needed because the number of points has changed. On the second check, the chosen points are $cp_1 = 376$ [see Fig. 4j] and $cp_2 = 526$ [see Fig. 4k]. Still another iteration is required because the distance from 397 to 376 is more than two observations. On the following iteration, the changepoints found are again $cp_1 = 376$ and $cp_2 = 526$, and hence the algorithm has converged.

In summary, we see for this example that the ICSS algorithm successfully eliminates the masking effect when there is more than one variance change. Specifically, the second changepoint $t = 526$ was not detected in the original $D_k(a[1 : 700])$ in Figure 1f, and the final $cp_1 = 376$ was appreciably away from the initial point $cp_1 = 342$ and closer to the true first changepoint, 391.

This algorithm can be included as part of the residual diagnostics for practitioners fitting time series models. Simulation results show that when we apply the ICSS algorithm to residuals of autoregressive processes, we obtain results similar to those obtained when we apply the ICSS algorithm to sequences of independent observations (see Sect. 4.5).

If the ICSS algorithm is programmed into a system that allows high resolution graphics, the user can look at the D_k plots as they are obtained. It is important to look at the plots, because a big outlier would create a significant peak in the D_k plot that might not be due to a variance change. In most cases it is easy to detect outliers affecting the D_k plot, because they will appear as sudden jumps; the slope of the D_k would not be changed. But it is advisable to complement the search for variance changes with a procedure for outlier detection (see Chang, Tiao, and Chen 1988).

3.1 Example: IBM Stock Prices

As an example, we apply the ICSS algorithm to a real data set. The series analyzed is the first difference of the logarithm of the IBM stock closing prices from May 17, 1961, to November 2, 1962, as reported by Box and Jenkins (1976). There is no substantial autocorrelation for this series.

The analysis using the ICSS algorithm converges in the first iteration with $k_1^* = 235$ and $k_2^* = 279$. Other authors have found similar results. Wichern, Miller, and Hsu fit an ARIMA(1, 1, 0) to the logarithm of the stock price, obtaining $\hat{\phi} = .13$ and the changepoints estimates at 180 and 235. Baufays and Rasson (1985) found two changes of variance at observations 235 and 280. Tsay (1988) found one variance change at 237. The results that we found using the ICSS algorithm agree with the maximum likelihood estimates of Baufays and Rasson. A Bayesian analysis that allows both the variance and the autoregressive parameter to change at the same points (Inclán 1991) indicates that there are both variance changes and parameter changes. The posterior modal values for the changepoints are the same points found by the ICSS algorithm.

The series analyzed in this section illustrates the use of cumulative sums of squares to diagnose the existence of changes in variance. Using the estimates of the changepoints obtained with the ICSS algorithm, we can estimate the variances for the different pieces of the series and use them to obtain a series with homogeneous variance. This "corrected series" should then be used to obtain other residual diagnostics, like the Q statistic, which would be distorted if calculated with the series that has variance changes (see Inclán 1991). The Fortran program of the ICSS algorithm is available from us upon request; send an electronic mail message to inclan@guvax.georgetown.edu.

Table 3. Standardization Formulas for the $LR_{n-1,n}$ Statistic, n = 1,2,3

$St\ LR_{0,1} = (LR_{0,1} - (5.01 - 14.5/\sqrt{T}))/2.15$
$St\ LR_{1,2} = (LR_{1,2} - (9.14 - 14.76/\sqrt{T} - \log(\hat{\Delta})(2.11 - .31(\hat{\kappa}/T - .5))))/2.17$
$St\ LR_{2,3} = (LR_{2,3} - (9.81 - 44.34/\sqrt{T}))/(2.41 - 6.82/\sqrt{T})$

NOTE: $LR_{1,2}$ requires the maximum likelihood estimators $\hat{\kappa}$ and $\hat{\Delta} = \hat{\tau}_1^2/\hat{\tau}_0^2$.

4. SIMULATION EXPERIMENT

The performance of different procedures used to determine the number of variance changes in a series can be measured in several ways; the main way is by the number of "correct identifications." Another aspect to consider is the computational requirements of each approach, which can be measured in terms of the CPU time. We now present the results of a simulation experiment to compare the performance of the ICSS algorithm with the likelihood ratio and a Bayesian approach.

The simulation experiment has two separate parts, the first for one changepoint and the second for two changepoints in the generated series. For each part, we used 1,000 replicates per design point, each of them a series of length T of independent $N(0, 1)$ random variables. The Gaussian random number generator is based on the modified polar method and uses uniform random numbers from the Bratley, Fox, and Schrage (1987) portable uniform random number generator. The routines are programmed in Fortran and were run in a Sparcstation 1+ at the Smeal College of Business of the Pennsylvania State University.

The same series is subject to each of the procedures considered to facilitate paired comparisons between procedures. For Part One, we used three series lengths (T = 100, 200, and 500), three locations of the changepoint ($\kappa_1 = .25T$, $.50T$, and $.75T$), and two values for the variance ratio (Δ = 2 and 3). For Part Two, we used the same three series lengths (T = 100, 200, and 500), three different sets of locations of the change points [$(\kappa_1, \kappa_2) = (.33T, .66T)$, $(.20T, .80T)$, and $(.58T, .80T)$], and six possible configurations of variances $(\tau_0^2, \tau_1^2, \tau_2^2)$, where we impose the condition that $\tau_0^2 = 1$ but keep the corresponding ratios $\Delta_1 = \tau_1^2/\tau_0^2$ and $\Delta_2 = \tau_2^2/\tau_0^2$. Some of the permutations of the variances were eliminated based on symmetry considerations; for example, a series with variances (1, 2, 4) is equivalent to a series with variances (4, 2, 1) taken in reverse order.

4.1. Assessing the Evidence with Respect to the Number of Change Points Using Likelihood Ratio Tests and the Posterior Odds Ratio

Before presenting the results of the simulation, we give some details about the procedures to detect multiple points of variance change using likelihood ratio tests and using posterior odds ratios.

Let $LR_{m,n}$ denote the likelihood ratio statistic for testing H_0: $N_T = m$ against H_a: $N_T = n$. The expres-

Table 4. Frequency Distribution of $\hat{N}_T$ Obtained with the ICSS Algorithm, Likelihood Ratio Tests and Log Posterior Odds for Series with One Variance Change

				ICSS algorithm			LR tests			Log posterior odds		
Δ	T	κ/T	$\hat{N}_T$	0	1	≥2	0	1	≥2	0	1	≥2
1	100			**97.1**	2.8	.1	**95.7**	3.9	.4	**89.3**	9.8	.9
	200			**96.5**	3.3	.2	**94.8**	4.4	.8	**92.2**	7.3	.5
	500			**95.5**	4.2	.3	**95.1**	4.4	.5	**94.7**	5.1	.2
2	100											
		.25		84.6	**14.8**	.6	73.5	**25.1**	1.4	44.6	**52.8**	2.6
		.50		48.4	**50.2**	1.4	58.9	**38.1**	3.0	36.8	**59.4**	3.8
		.75		61.4	**37.3**	1.3	68.6	**28.6**	2.8	51.3	**45.0**	3.7
	200											
		.25		55.9	**41.3**	2.8	43.2	**52.9**	3.9	25.3	**71.1**	3.6
		.50		14.4	**82.2**	3.4	25.2	**70.2**	4.6	13.2	**83.4**	3.4
		.75		27.1	**70.1**	2.8	37.6	**58.5**	3.9	28.6	**68.4**	3.0
	500											
		.25		5.5	**89.1**	5.4	4.3	**91.1**	4.6	2.6	**94.5**	2.9
		.50		0	**95.0**	5.0	.7	**94.6**	4.7	.5	**96.4**	3.1
		.75		2.1	**93.3**	4.6	3.9	**91.8**	4.3	2.9	**93.2**	3.9
3	100											
		.25		59.2	**39.8**	1.0	34.1	**61.6**	4.3	12.5	**82.5**	5.0
		.50		7.5	**89.4**	3.1	13.2	**81.0**	5.8	4.2	**87.0**	8.8
		.75		17.2	**80.5**	2.3	25.4	**68.7**	5.9	13.6	**77.7**	8.7
	200											
		.25		10.2	**85.7**	4.1	5.2	**90.5**	4.3	1.5	**94.5**	4.0
		.50		.3	**93.1**	6.6	.5	**94.1**	5.4	.3	**91.4**	8.3
		.75		1.6	**94.9**	3.5	3.1	**92.9**	4.0	1.6	**91.1**	7.3
	500											
		.25		0	**93.9**	6.1	0	**95.8**	4.2	0	**95.6**	4.4
		.50		0	**92.4**	7.6	0	**94.7**	5.3	0	**95.3**	4.7
		.75		0	**93.8**	6.2	0	**95.4**	4.6	0	**95.1**	4.9

NOTE: For series with Δ = 1, there are 10,000 replicates for the ICSS algorithm; all other cases have 1,000 replicates.

Table 5. Frequency Distribution of $\hat{N}_T$ Obtained with the ICSS Algorithm, Likelihood Ratio (LR) Tests and Log Posterior Odds for Series with Two Variance Changes

		$(\kappa_1/T, \kappa_2/T)$	*ICSS algorithm*			*LR tests*			*Log posterior odds*		
(Δ_1, Δ_2)	T	$\hat{N}_T$	*0*	*1*	*2*	*0*	*1*	*2*	*0*	*1*	*2*
(2, 4)	100	(.33, .67)	9.4	88.4	**2.0**	12.6	84.0	**3.1**	2.4	89.3	**8.1**
	200	(.33, .67)	.2	76.1	**20.8**	.4	92.1	**7.1**	0	74.0	**23.0**
	500	(.33, .67)	0	9.4	**84.3**	0	35.2	**59.4**			
	100	(.20, .80)	42.7	57.0	**0.2**	38.9	58.9	**2.1**	15.6	78.3	**5.8**
	200	(.20, .80)	11.0	81.5	**6.7**	8.0	85.5	**6.2**	3.0	70.0	**24.0**
	500	(.20, .80)	0	25.7	**69.3**	0	42.3	**52.6**			
	100	(.58, .80)	7.5	90.3	**2.2**	12.4	84.3	**3.2**	4.9	88.1	**7.0**
	200	(.58, .80)	0.3	90.7	**8.7**	.5	94.8	**4.6**	0	89.0	**11.0**
	500	(.58, .80)	0	35.6	**59.5**	0	65.2	**31.5**			
(4, 2)	100	(.33, .67)	38.8	34.5	**26.0**	25.3	60.2	**12.6**	9.3	50.6	**38.0**
	200	(.33, .67)	1.5	28.2	**67.0**	1.5	58.3	**37.5**	1.0	33.0	**62.0**
	500	(.33, .67)	0	1.7	**92.4**	0	10.0	**86.1**			
	100	(.20, .80)	86.9	8.5	**4.5**	31.9	59.9	**7.4**	8.6	54.8	**36.3**
	200	(.20, .80)	29.9	38.4	**30.4**	3.3	67.9	**27.2**	0	39.0	**60.0**
	500	(.20, .80)	0	11.3	**84.8**	0	18.2	**78.1**			
	100	(.58, .80)	9.8	70.0	**19.6**	16.6	73.9	**9.1**	7.0	64.1	**27.7**
	200	(.58, .80)	0.3	51.6	**46.7**	1.2	73.9	**24.0**	0	60.0	**37.0**
	500	(.58, .80)	0	11.7	**83.1**	0	33.8	**62.8**			
(0.5, 2)	100	(.33, .67)	21.4	53.0	**24.8**	34.0	55.0	**9.8**	14.7	44.1	**39.7**
	200	(.33, .67)	2.2	33.1	**60.8**	6.1	56.9	**34.4**	1.0	33.0	**65.0**
	500	(.33, .67)	0	3.4	**89.4**	0	11.6	**83.1**			
	100	(.20, .80)	24.8	47.8	**27.2**	29.6	58.6	**10.8**	15.6	46.3	**37.1**
	200	(.20, .80)	2.8	37.0	**57.8**	3.5	58.1	**36.4**	4.0	31.0	**63.0**
	500	(.20, .80)	0	3.6	**90.2**	0	12.9	**82.3**			
	100	(.58, .80)	61.0	34.7	**4.2**	63.6	31.5	**4.2**	47.7	30.2	**20.8**
	200	(.58, .80)	27.8	42.0	**28.2**	30.4	47.8	**20.3**	17.0	21.0	**59.0**
	500	(.58, .80)	1.3	6.4	**84.1**	1.7	14.8	**79.4**			
(.5, .25)	100	(.58, .80)	20.2	78.1	**1.5**	15.8	81.3	**2.6**	2.4	91.7	**5.7**
	200	(.58, .80)	1.4	91.8	**5.8**	.7	94.9	**4.0**	0	85.0	**15.0**
	500	(.58, .80)	0	40.6	**55.6**	0	68.8	**29.6**			
(2, .5)	100	(.58, .80)	90.0	2.7	**7.1**	61.7	28.9	**8.4**	34.1	39.1	**25.1**
	200	(.58, .80)	67.5	5.4	**26.3**	25.2	40.9	**32.3**	17.0	30.0	**51.0**
	500	(.58, .80)	1.6	2.4	**89.8**	.2	10.1	**86.0**			
(.25, .5)	100	(.58, .80)	23.6	67.1	**9.0**	33.1	60.3	**6.0**	13.9	63.4	**22.2**
	200	(.58, .80)	1.3	69.0	**28.2**	3.5	75.7	**20.0**	3.0	55.0	**39.0**
	500	(.58, .80)	0	28.0	**65.8**	0	35.4	**61.8**			

sion for the special case $LR_{0,1}$ is given in (5). (For the general case, see Inclán 1991). A strategy to determine the number of changepoints is to consider $LR_{n-1,n}$, for $n = 1, 2, \ldots$. That is, if there is evidence in the data to reject $N_T = n - 1$, then consider one more change.

The critical values for the tests must be obtained by simulation, because this is not a regular problem. In the case of changepoints in the parameter of the exponential distribution, Haccou and Meelis (1988) showed that the asymptotic null distribution of the likelihood ratio test for one change against no change is an extreme value distribution. The simulation results of Inclán (1991) indicate that this asymptotic distribution may also apply to $LR_{m,n}$ here, and that we can standardize the $LR_{n-1,n}$ statistic according to the formulas in Table 3 for reference to the extreme value distribution with distribution function $F(x) = \exp(-e^{-x})$.

The posterior odds for assessing the evidence in the data about the number of change points are

$$
\begin{aligned}
K_{n,m} &= \frac{p(N_T = n \mid \mathbf{a})}{p(N_T = m \mid \mathbf{a})} = \left(\frac{\lambda}{1-\lambda}\right)^{n-m} \\
&\quad \times \frac{m!(T-1-m)! \sum_{\kappa_1} \cdots \sum_{\kappa_n} p(\mathbf{a}, \kappa \mid N_T = n)}{n!(T-1-n)! \sum_{\kappa_1} \cdots \sum_{\kappa_m} p(\mathbf{a}, \kappa \mid N_T = m)} \\
&= \left(\frac{\lambda}{1-\lambda}\right)^{n-m} \left(\frac{c^{\nu/2}}{\Gamma(\nu/2)}\right)^{n-m} \\
&\quad \times \frac{\sum_{\kappa_1} \cdots \sum_{\kappa_n} \prod_{j=0}^{n} \left\{ \Gamma\left(\frac{d_j+\nu}{2}\right) \left(\frac{1}{2} \sum_{t=\kappa_j+1}^{\kappa_{j+1}} a_t^2 + c\right)^{-(d_j+\nu)/2} \right\}}{\sum_{\kappa_1} \cdots \sum_{\kappa_m} \prod_{j=0}^{m} \left\{ \Gamma\left(\frac{d_j+\nu}{2}\right) \left(\frac{1}{2} \sum_{t=\kappa_j+1}^{k_{j+1}} a_t^2 + c\right)^{-(d_j+\nu)/2} \right\}},
\end{aligned} \tag{15}
$$

where $p(N_T = n \mid \mathbf{a})$ is the posterior probability that the series has n changes of variance; $p(\mathbf{a}, \kappa \mid N_T)$ is the conditional probability for $\mathbf{a} = (a_1, \ldots, a_T)'$ and $\kappa = (\kappa_1, \ldots, \kappa_{N_T})'$ given the number of change points N_T; λ is the prior probability of having a change in variance at each point in time;

Table 6. Summaries from the Sampling Distribution of k^* for Series with One Variance Change

Δ	T	κ	Mode of k^*	Mean of k^*	SD of k^*
2	100				
		25	25	43.40	17.25
		50	50	54.83	11.84
		75	75	68.08	15.81
	200				
		50	50	75.10	29.24
		100	101	107.60	15.38
		150	150	142.98	22.19
	500				
		125	125	161.31	48.90
		250	250	259.75	19.99
		375	375	367.39	26.47
3	100				
		25	25	39.78	15.23
		50	50	54.86	7.49
		75	75	72.95	9.95
	200				
		50	50	69.03	23.05
		100	100	105.94	9.31
		150	150	149.33	9.10
	500				
		125	125	149.59	32.58
		250	250	256.65	11.30
		375	375	374.36	10.55

NOTE: There are 1,000 replicates for each row of the table.

and c and ν are the hyperparameters of the inverted gamma used prior for the τ_i, $i = 0, \ldots, N_T$ (see Inclán 1991).

A systematic way of using the posterior odds to determine the number of variance changes on a given series is to calculate $K_{n,n-1}$ for $n = 1, 2, \ldots$. A decision can be reached without specifying a loss function if a maximum number of changepoints $N_{\max}$ is specified. Calculate $K_{n,n-1}$ for $n = 1, \ldots, N_{\max}$; from these odds, obtain $p(N_T \mid \mathbf{a}, N_T \leq N_{\max})$ and choose the value of N_T for which this conditional probability is maximum.

4.2. Results of Part One: Series With One Changepoint

Table 4 gives the frequency distribution of $\hat{N}_T$ obtained with the ICSS algorithm, with likelihood ratio tests and with the log posterior odds. The column corresponding to the "correct identifications" has been highlighted. For the likelihood ratio procedure, we follow a sequence of likelihood ratio tests up to a maximum value of two changepoints.

We conclude that for series of 100 observations, it is hard to detect a small variance change (variance ratio $\Delta = 2$)—particularly when it appears at the beginning of the series and we are using the ICSS algorithm. Once we have 200 observations or the variance ratio is larger ($\Delta = 3$), the correct identifications occur more than 80% of the time if the changepoint is in the middle of the series. The ICSS algorithm performance improves notably as we have longer series and larger variance ratios. The Bayesian procedure is the best for $\Delta = 2$ and gives comparable results to the other approaches with $\Delta = 3$.

4.3. Results of Part Two: Series With Two Changepoints

The frequency distribution of $\hat{N}_T$ is presented in Table 5, with the column corresponding to the correct identifications in bold type. For series of 100 observations, we find only one changepoint most of the time. The most difficult situation is when the variances change in a monotone way; that is, the variance increases at the first change point and increases again at the second change point. If this is the case, then it is necessary to have a large number of observations (e.g., 500), to be able to get $\hat{N}_T = 2$ more than half of the time.

In almost every instance with 200 observations or more, the ICSS algorithm gives better results than the likelihood ratio tests. The best results across location of the points of change are at $(\kappa_1/T, \kappa_2/T) = (.33, .67)$, when the intervals between changepoints are of equal length. Across the six possible permutations of variances, the best results are found mostly when the large variance is in the middle, $(\Delta_1, \Delta_2) = (4, 2)$.

The posterior odds ratio was not obtained for series with $T = 500$, because it would have required a total of 1,358 CPU hours. In fact, the runs with $T = 200$ for the log posterior odds have only 100 replicates instead of the 1,000 in the rest of the runs in the experiment.

4.4 Sampling Distribution of k^*

The same simulation provides information about the sampling distribution of k^*, the point where $\max_k |D_k(a[1:T])|$ is attained. This distribution is skewed; k^* is biased towards the middle of the series. What makes the ICSS algorithm work well is that the mode of k^* is exactly at the point where the change in variance occurs. The values of k^* become increasingly concentrated around the true changepoint as the sample size increases or as the variance ratio increases.

Tables 6 and 7 present summaries of this distribution. Table 6 corresponds to series with one changepoint. Table 7 presents the results for series with two changepoints, for which the mode of k^* is attained at the change point where $E[D_k]$ is largest, in absolute value (see Fig. 3).

4.5 The ICSS Algorithm Applied to Residuals

The same series used in the first part of the simulation experiment were used to obtain an autoregressive process of order 1: $Y_t = \phi Y_{t-1} + a_t$. Then the ICSS algorithm was applied to the residuals $\hat{a}_t = Y_t - \hat{\phi} Y_{t-1}$, where $\hat{\phi}$ is the ordinary least squares estimate of ϕ. Table 8 shows the results for two values of ϕ, the percentage of series for which $\hat{N}_T = 1$ is very close to the figures presented in Table 4. Further research is underway and will be reported elsewhere.

4.6 CPU Time Requirements

One of the concerns that motivated the design of the ICSS algorithm is the heavy computational burden involved in the calculation of posterior odds. When using likelihood ratio tests, the burden is partially alleviated by obtaining the maximum of the log-likelihood function using the Baufays and Rasson (1985) algorithm. Nonetheless, it is still a heavy computation when dealing with series of hundreds of observations and several changepoints. The ICSS algorithm avoids calculating a function at all possible locations of the change-

Table 7. Summaries from the Sampling Distribution of k^* for Series With Two Variance Changes

(Δ_1, Δ_2)	T	(κ_1, κ_2)	Mode of k^*	Mean of k^*	SD of k^*
(2, 4)	100	(33, **67**)	67	61.05	12.34
	200	(66, **134**)	134	122.54	21.23
	500	(165, **335**)	335	314.19	37.57
	100	(20, **80**)	80	63.41	19.41
	200	(40, **160**)	160	131.30	35.82
	500	(100, **400**)	401	348.11	71.34
	100	(58, **80**)	58	68.14	9.91
	200	(**116,** 160)	116	133.18	16.36
	500	(**290,** 400)	292	329.10	37.42
(4, 2)	100	(**33,** 67)	33	37.23	7.27
	200	(**66,** 134)	66	70.43	7.32
	500	(**165,** 335)	165	169.14	6.87
	100	(**20,** 80)	20	32.61	15.51
	200	(**40,** 160)	40	53.70	21.45
	500	(**100,** 400)	100	113.09	20.31
	100	(**58,** 80)	58	59.14	5.05
	200	(**116,** 160)	116	117.29	5.14
	500	(**290,** 400)	290	291.60	5.78
(.5, 2)	100	(33, **67**)	67	68.97	8.43
	200	(66, **134**)	134	137.22	8.00
	500	(165, **335**)	335	339.32	6.96
	100	(20, **80**)	80	76.35	16.20
	200	(40, **160**)	160	158.89	16.25
	500	(100, **400**)	400	400.86	10.98
	100	(58, **80**)	80	75.06	17.00
	200	(116, **160**)	160	157.51	22.39
	500	(290, **400**)	400	402.67	13.12
(.5, .25)	100	(**58,** 80)	58	52.35	10.94
	200	(**116,** 160)	116	109.94	15.42
	500	(**290,** 400)	289	281.13	20.03
(2, .5)	100	(58, **80**)	80	64.10	16.99
	200	(116, **160**)	160	140.47	26.20
	500	(290, **400**)	400	380.02	42.26
(.25, .5)	100	(**58,** 80)	58	50.25	9.50
	200	(**116,** 160)	115	106.63	11.95
	500	(**290,** 400)	290	277.41	17.75

Table 8. Percentage of $\hat{N}_T = 1$ Obtained with the ICSS Algorithm for Series With One Variance Change, Residuals from AR(1)

				$\phi = .6$			$\phi = .9$		
Δ	T	κ/T	$\hat{N}_T$	0	1	≥2	0	1	≥2
1	100			**97.2**	2.7	.8	**97.2**	2.7	.1
	200			**96.5**	3.2	.2	**96.7**	3.1	.2
	500			**95.6**	4.2	.2	**95.6**	4.1	.3
2	100	.25		86.4	**13.1**	.5	85.4	**13.9**	.7
		.50		54.0	**44.3**	1.7	52.5	**45.7**	1.8
		.75		61.0	**37.5**	1.5	58.3	**40.9**	.8
	200	.25		61.3	**36.4**	2.2	58.0	**40.5**	1.5
		.50		15.4	**81.2**	3.3	14.3	**81.9**	3.8
		.75		31.5	**65.7**	2.8	33.0	**63.1**	3.9
	500	.25		6.0	**88.3**	5.7	6.0	**88.7**	5.3
		.50		.2	**94.5**	5.3	.1	**94.0**	5.9
		.75		1.5	**93.2**	5.3	2.0	**91.7**	6.3
3	100	.25		68.4	**30.8**	.8	68.5	**29.9**	1.4
		.50		11.3	**86.1**	2.6	11.8	**84.8**	3.4
		.75		21.0	**75.9**	3.1	20.5	**76.3**	3.2
	200	.25		10.9	**85.4**	3.7	12.9	**83.3**	3.8
		.50		.1	**95.8**	4.1	.1	**94.8**	5.1
		.75		2.1	**94.1**	3.8	.9	**95.2**	3.9
	500	.25		0	**93.0**	7.0	0	**95.1**	4.9
		.50		0	**91.5**	8.5	0	**94.0**	6.0
		.75		0	**93.7**	6.3	0	**94.0**	6.0

NOTE: For series with $\Delta = 1$, there are 10,000 replicates for the ICSS algorithm; all other cases have 1,000 replicates.

points by looking for the changepoints in an iterative manner. On average, after cutting and analyzing the pieces, we need to perform $O(T)$ operations. This section presents actual CPU times obtained in the simulations described earlier.

The results can be summarized in the following regression models, estimated by least squares. Let CPU^m represent the average CPU time (in seconds) taken by each method: $m = B$ is posterior odds, $m = L$ is maximum likelihood, and $m = D$ is for the ICSS algorithm using D_k. Define three indicator variables: $X_1 = 1$ if $N_T = 1$, $X_2 = 1$ if $N_T = 2$, and $X_3 = 1$ if $N_T = 3$.

$$\begin{aligned}\log(\mathrm{CPU}^B) = &\underset{(.04)}{-12.14} + \underset{(.01)}{1.85}\log(T) - \underset{(.06)}{.35}X_2 \\ &- \underset{(.10)}{1.46}X_3 + \underset{(.01)}{.95}X_2\log(T) + \underset{(.02)}{1.94}X_3\log(T)\end{aligned}$$

$$s = .032 \qquad R^2 = 99.9\%$$

$$\begin{aligned}\log(\mathrm{CPU}^L) = &\underset{(.15)}{-8.66} + \underset{(.03)}{.8}\log(T) - \underset{(.2)}{2.48}X_2 \\ &- \underset{(.2)}{1.81}X_3 + \underset{(.04)}{1.16}(X_2 + X_3)\log(T)\end{aligned}$$

$$s = .147 \qquad R^2 = 99.4\%$$

$$\begin{aligned}\log(\mathrm{CPU}^D) = &\underset{(.06)}{-8.75} + \underset{(.01)}{.70}\log(T) \\ &+ \underset{(.02)}{1.15}X_1 + \underset{(.02)}{1.51}X_2 + \underset{(.02)}{1.74}X_3\end{aligned}$$

$$s = .105 \qquad R^2 = 98.6\%$$

The coefficients of $\log(T)$ indicate the order of magnitude of the CPU time required. The equation for CPU^B implies that the order of magnitude increases for larger N_T. In the case of the likelihood ratio approach, the CPU time is of the same order for $N_T = 2$ and $N_T = 3$, roughly $O(T^2)$. The model for the CPU time of the ICSS algorithm shows that the dependence on the series length is the same for all values of $\hat{N}_T$. This result clearly shows the different costs of using each of the methods presented here. It and the simulation results in Tables 4–8, lend support to the recommendation to use the ICSS algorithm when we need to analyze long series with multiple change points.

APPENDIX A: APPROXIMATE EXPECTED VALUE OF D_k

Let $\{a_t\}$ be a sequence of independent Normal random variables, with mean 0 and variance σ_t^2. For a fixed value of k, take a second-order Taylor expansion of the ratio C_k/C_T about the value $(E[C_k], E[C_T])$ to get the expected value of D_k as

$$E\left[\frac{C_k}{C_t}\right] - \frac{k}{T} = \frac{E[C_k]}{E[C_T]} - \frac{E[C_kC_T]}{E[C_T]^2} + \frac{E[C_k]E[C_T^2]}{E[C_T]^3} - \frac{k}{T} + o(T^{-1}), \quad \text{(A.1)}$$

where

$$E[C_k] = \sum_{t=1}^{k} \sigma_t^2 \qquad k = 1, \ldots, T \quad \text{(A.2)}$$

$$E[C_kC_T] = 3\sum_{t=1}^{k} \sigma_t^4 + \sum_{t=1}^{k}\sum_{s=t+1}^{T} \sigma_t^2\sigma_s^2 + \sum_{t=2}^{k}\sum_{s=1}^{t-1} \sigma_t^2\sigma_s^2 \quad \text{(A.3)}$$

We evaluate expressions (A.2) and (A.3) for each of three cases:

1. The series has homogeneous variance, $N_T = 0$, $\sigma_t^2 = \sigma^2$, for all t.
2. The series has one changepoint, $N_T = 1$ at κ_1, $\sigma_t^2 = \tau_0^2$ for $t = 1, \ldots, \kappa_1$ and $\sigma_t^2 = \tau_1^2$ for $t = \kappa_1 + 1, \ldots, T$.
3. There are two changepoints, $N_T = 2$, located at κ_1 and κ_2. The corresponding variances are τ_0^2 for $t = 1, \ldots, \kappa_1$, τ_1^2 for $t = \kappa_1 + 1, \ldots, \kappa_2$, and τ_2^2 for $t = \kappa_2 + 1, \ldots, T$.

Under the variance homogeneity, $N_T = 0$, $E[C_k] = k\sigma^2$, $E[C_kC_T] = k(T + 2)\sigma^4$, and $E[D_k] \approx 0$.

When there is one variance change at κ_1,

$$\begin{aligned}E[C_k] &= k\tau_0^2 && 1 \le k \le \kappa_1 \\ &= \kappa_1\tau_0^2 + (k - \kappa_1)\tau_1^2 && \kappa_1 \le k \le T\end{aligned} \quad \text{(A.4)}$$

and

$$\begin{aligned}E[C_kC_T] &= k(\kappa_1 + 2)\tau_0^4 + k(T - \kappa_1)\tau_0^2\tau_1^2 && 1 \le k \le \kappa_1 \\ &= \kappa_1(\kappa_1 + 2)\tau_0^4 + (k - \kappa_1)(T - \kappa_1 + 2)\tau_1^4 \\ &\quad + \kappa_1(T + k - 2\kappa_1)\tau_0^2\tau_1^2 && \kappa_1 \le k \le T.\end{aligned} \quad \text{(A.5)}$$

Hence, in terms of the ratio of variances $\Delta_1 = \tau_1^2/\tau_0^2$,

$$\begin{aligned}E[D_k] &\approx \frac{k(\kappa_1^2 + (T - \kappa_1)(T - \kappa_1 + 2)\Delta_1^2 + 2(\kappa_1 - 1)(T - \kappa_1)\Delta_1)}{(\kappa_1 + (T - \kappa_1)\Delta_1)^3} - \frac{k}{T} \qquad 1 \le k \le \kappa_1 \\ &\approx \frac{\kappa_1^3 + \kappa_1((\kappa_1 + 2)(k - \kappa_1) + 2(T - \kappa_1)(2\kappa_1 + 1))\Delta_1}{(\kappa_1 + (T - \kappa_1)\Delta_1)^3} \\ &\quad + \frac{\kappa_1((T - \kappa_1)(T - \kappa_1 + 2) + 2(k - \kappa_1)(2(T - \kappa) + 1))\Delta_1^2 + (T - \kappa_1)^2(k - \kappa_1)\Delta_1^3}{(\kappa_1 + (T - \kappa_1)\Delta_1)^3} - \frac{k}{T} \qquad \kappa_1 \le k \le T.\end{aligned} \quad \text{(A.6)}$$

In the case of two variance changes, the second-order approximation to $E[D_k]$ can be evaluated using

$$\begin{aligned}E[C_k] &= k\tau_0^2 && 1 \le k \le \kappa_1 \\ &= \kappa_1\tau_0^2 + (k - \kappa_1)\tau_1^2 && \kappa_1 \le k \le \kappa_2 \\ &= \kappa_1\tau_0^2 + (\kappa_2 - \kappa_1)\tau_1^2 + (k - \kappa_2)\tau_2^2 && \kappa_2 \le k \le T\end{aligned} \quad \text{(A.7)}$$

and

$$
\begin{aligned}
E[C_kC_T] &= k(\kappa_1+2)\tau_0^4 + k(\kappa_2-\kappa_1)\tau_0^2\tau_1^2 + k(T-\kappa_2)\tau_0^2\tau_2^2 \qquad 1 \le k \le \kappa_1 \\
&= \kappa_1(\kappa_1+2)\tau_0^4 + (k-\kappa_1)(\kappa_2-\kappa_1)\tau_1^4 + \kappa_1(\kappa_2+k-2\kappa_1)\tau_0^2\tau_1^2 + \kappa_1(T-\kappa_2)\tau_0^2\tau_2^2 + (k-\kappa_1)(T-\kappa_2)\tau_1^2\tau_2^2 \qquad \kappa_1 \le k \le \kappa_2 \\
&= \kappa_1(\kappa_1+2)\tau_0^4 + (\kappa_2-\kappa_1)(\kappa_2-\kappa_1+2)\tau_1^4 + (k-\kappa_2)(T-\kappa_2+2)\tau_2^4 + 2\kappa_1(\kappa_2+k-2\kappa_1)\tau_0^2\tau_1^2 \\
&\quad + \kappa_1(T+k-2\kappa_2)\tau_0^2\tau_2^2 + (\kappa_2-\kappa_1)(T+k-2\kappa_2)\tau_1^2\tau_2^2 \qquad \kappa_2 \le k \le T
\end{aligned} \tag{A.8}
$$

APPENDIX B: PROOF OF THEOREM 1

Let $\xi_t = a_t^2 - \sigma_a^2$, so $E[\xi_t] = 0$ and $\sigma^2 = \text{var}(\xi_i) = 2\sigma_a^2$. Let

$$X_n(t) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1},$$

where $S_n = \xi_1 + \cdots + \xi_n$. By Donsker's theorem (Billingsley 1968, thm. 10.1), $X_n \xrightarrow{\mathcal{D}} W$, so $\{X_n(t) - tX_n(1)\} \xrightarrow{\mathcal{D}} W^0$ (Billingsley 1968, thm. 5.1). Let $nt = k$, $k = 1, 2, \ldots, n$. Then

$$
\begin{aligned}
X_n(t) - tX_n(1) &= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{n}} S_{[nt]} - \frac{t}{\sqrt{2}\sigma_a^2\sqrt{n}} S_{[n]} + \frac{(nt-[nt])}{\sqrt{2}\sigma_a^2\sqrt{n}} \xi_{[nt]+1} \\
&= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{n}} \left[S_k - \frac{k}{n} S_n \right] + \frac{(nt-[nt])}{\sqrt{2}\sigma_a^2\sqrt{n}} \xi_{[nt]+1} \\
&= \frac{1}{\sqrt{2}\sigma_a^2\sqrt{n}} \left[\sum_{t=1}^{k} (a_t^2 - \sigma_a^2) - \frac{k}{n} \sum_{t=1}^{n} (a_t^2 - \sigma_a^2) \right] + \frac{(nt-[nt])}{\sqrt{2}\sigma_a^2\sqrt{n}} \xi_{[nt]+1}.
\end{aligned}
$$

Note that

$$\left[\sum_{t=1}^{k} (a_t^2 - \sigma_a^2) - \frac{k}{n} \sum_{i=1}^{n} (a_i^2 - \sigma_a^2) \right] = \left(\sum_{t=1}^{n} a_t^2 \right) D_k.$$

Thus

$$\sqrt{\frac{n}{2}}\, D_k \left(\frac{\frac{1}{n} \sum_{i=1}^{n} a^2}{\sigma_a^2} \right) = X_n(t) - tX_n(1) - \frac{(nt-[nt])}{\sqrt{2}\sigma_a^2\sqrt{n}} \xi_{[nt]+1}.$$

As $n \to \infty$,

$$X_n(t) - tX_n(1) \xrightarrow{\mathcal{D}} W^0$$

and

$$\sup_t \left| \frac{(nt-[nt])}{\sqrt{2}\sigma_a^2\sqrt{n}} \xi_{[nt]+1} \right| \xrightarrow{P} 0.$$

Therefore (Billingsley 1968, thm. 4.1, p. 25),

$$\sqrt{\frac{n}{2}}\, D_k \left(\frac{\frac{1}{n} \sum_{i=1}^{n} a_i^2}{\sigma_a^2} \right) \xrightarrow{\mathcal{D}} W^0.$$

By the law of large numbers, $(1/n) \sum_{i=1}^{n} a_i^2 \to \sigma_a^2$, so $\sqrt{(n/2)}\, D_k \xrightarrow{\mathcal{D}} W^0$.

The distribution of $\sup_t |W_t^0|$ was given in equation (11.39) of Billingsley (1968):

$$P\left\{ \sup_t |W_t^0| \le b \right\} = 1 + 2 \sum_{k=1}^{\infty} (-1)^k e^{-2k^2b^2} \qquad b > 0.$$

[Received June 1992. Revised September 1993.]

REFERENCES

Abraham, B., and Wei, W. W. S. (1984), "Inferences About the Parameters of a Time Series Model With Changing Variance," *Metrika*, 31, 183–194.

Baufays, P., and Rasson, J. P. (1985), "Variance Changes in Autoregressive Models," in *Time Series Analysis: Theory and Practice 7*, ed. O. D. Anderson, New York: North-Holland, pp. 119–127.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: John Wiley.

Booth, N. B., and Smith, A. F. M. (1982), "A Bayesian Approach to Retrospective Identification of Change-Points," *Journal of Econometrics*, 19, 7–22.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden Day.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987), *A Guide to Simulation* (2nd ed.), New York: Springer-Verlag.

Broemeling, L. D., and Tsurumi, H. (1987), *Econometrics and Structural Change*, New York: Marcel-Dekker.

Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time," *Journal of the Royal Statistical Society*, Ser. B, 37, 149–163.

Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.

Cobb, G. W. (1978), "The Problem of the Nile: Conditional Solution to a Change Point Problem," *Biometrika*, 65, 243–251.

Commenges, D., Seal, J., and Pinatel, F. (1986), "Inference About a Change Point in Experimental Neurophysiology," *Mathematical Biosciences*, 80, 81–108.

Haccou, P., and Meelis, E. (1988), "Testing for the Number of Change Points in a Sequence of Exponential Random Variables," *Journal of Statistical Computing Simulation*, 30, 285–298.

Haccou, P., Meelis, E., and van de Geer, S. (1988), "The Likelihood Ratio Test of the Change Point Problem for Exponentially Distributed Random Variables," *Stochastic Processes and Their Applications*, 27, 121–139.

Hinkley, D. V. (1971), "Inference About the Change Point From Cumulative Sum Test," *Biometrika*, 58, 509–523.

Hsu, D. A. (1977), "Tests for Variance Shift at an Unknown Time Point," *Applied Statistics*, 26, 279–284.

——— (1979), "Detecting Shifts of Parameter in Gamma Sequences, With Applications to Stock Price and Air Traffic Flow Analysis," *Journal of the American Statistical Association*, 74, 31–40.

——— (1982), "A Bayesian Robust Detection of Shift in the Risk Structure of Stock Market Returns," *Journal of the American Statistical Association*, 77, 29–39.

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

Inclán, C. (1991), "Retrospective Detection of Sudden Changes of Variance in Time Series," Unpublished Ph.D. thesis, University of Chicago, Dept. of Statistics.

Menzefricke, U. (1981), "A Bayesian Analysis of a Change in the Precision of a Sequence of Independent Normal Random Variables at an Unknown Time Point," *Applied Statistics*, 30, 141–146.

Smith, A. F. M. (1975), "A Bayesian Approach to Inference About a Change-Point in a Sequence of Random Variable," *Biometrika*, 62, 407–416.

——— (1980), "Change-Point Problems: Approaches and Applications," in *Bayesian Statistics*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Valencia: Valencia University Press, pp. 83–98.

Tsay, R. (1988), "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, 7, 1–20.

Tyssedal, J. S., and Tjøstheim, D. (1988), "An Autoregressive Model With Suddenly Changing Parameters and an Application to Stock Market Prices," *Applied Statistics*, 37, 353–369.

Wichern, D. W., Miller, R. B., and Hsu, D. A. (1976), "Changes of Variance in First Order Autoregressive Time Series Models—With an Application," *Applied Statistics*, 25, 248–256.

Worsley, K. J. (1986), "Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables," *Biometrika*, 73, 91–104.