UvA ECONOMETRICS

Discussion Paper: 2005/06

Detecting change-points in multidimensional stochastic processes

Jan G. De Gooijer

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Detecting Change-points in Multidimensional Stochastic Processes

Jan G. De Gooijer

October 27, 2005

Department of Quantitative Economics
University of Amsterdam
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
Telephone: +31–20–525 4244; Fax: +31–20–525 4349
e-mail: j.g.degooijer@uva.nl

Abstract

A general test statistic for detecting change-points in multidimensional stochastic processes with unknown parameters is proposed. The test statistic is specialised to the case of detecting changes in sequences of covariance matrices. Large-sample distributional results are presented for the test statistic under the null hypothesis of no-change. Tables of selected quantiles of the proposed test are also given. The finite-sample properties of the test statistic are compared with two other test statistics proposed in the literature. Using a binary segmentation procedure, the potential of the various test statistics is investigated in a multidimensional setting both via simulations and the analysis of a real life example. In general, all test statistics become more effective as the dimension increases, avoiding the determination of too many "incorrect" change-point locations in a one-dimensional setting.

Key Words: Binary segmentation; Cramér-von Mises; Covariance matrices; Cumulative sum; Likelihood ratio.

1 Introduction

Procedures for detecting changes of variances in an ordered sequence of (possibly independent) observations taken from a multidimensional stochastic process can help to elucidate the structure of the process. For instance, it is well known that homogeneity of variance in a time series of observations taken from a single financial risk factor does not necessarily imply a homogeneous behaviour in variance of all possible risk factors simultaneously. Hence, a univariate change-point detection procedure may well fail to reject the assumption of constant variance underlying the model fitted to each individual series. The determination of arrival times for various phases on regional seismograms is another important change-point problem. Isolating onset times of seismic phases is usually done by experienced analysts at a single channel. The development of large digital seismic networks, requires the availability of some sort of automatic way of sequentially examining phases coming from multiple channels.

The case of abrupt changes in some of the parameters of a statistical model has been an area of interest for many years. In view of the limited amount of space a number of key references are given below. However, before doing so, some notation is needed to describe the change-point problem more precisely. Let X be a random variable whose probability density function is characterized by a parameter θ, i.e. $X \sim f(x|\theta)$. Then consider n consecutively observed independent random variables $\{X_1, X_2, \ldots, X_n\}$, where the density of X_i is $f(x|\theta_i)$. Let θ_0 be the initial starting value which may or may not be specified. The problem of change-point detection in a univariate setting is that of testing the null hypothesis $\theta_i = \theta_0$ $(i = 1, 2, \ldots, n)$ against the alternative hypothesis $\theta_i = \theta_0$ (i=1,2,...,m) and $\theta_i = \theta_0 + \delta$ $(i = m+1, m+2, \ldots, n)$, where the change-point m is not specified and δ is the amount of change in the parameters.

Broadly speaking the detection of a change-point has been dealt with in three different contexts, two of which are parametric. The parametric approaches are the 'Bayes-type' and the maximum likelihood (ML) approach. Chernoff and Zacks (1964) first introduced a Bayes-type statistic for the detection of a change in the mean at unknown times in a sequence of independent normal random variables against a one-sided alternative. This methodology uses the Bayesian approach to eliminate nuisance parameters and then derives the unconditional likelihood ratio statistic. The work was extended by Gardner (1969) to detect the change-point against two-sided alternatives. Sen and Srivastava (1973) then considered mean changes in a sequence of multivariate normal random variables. Kander and Zacks (1966) examined the one-sided change in the one-parameter exponential class of distributions and MacNeill (1978) extended the technique to the two-sided change.

Others have used a ML ratio approach to develop statistics for the detection of parameter changes at unknown points. Among them are Quandt (1958, 1960), Hinkley (1971), Sen and Srivastava (1975), Hawkins (1977), Worsley (1979), Esterby and El-Shaarawi (1981), and Horváth (1993). Sen and Srivastava (1975) compared the Bayes-type statistics against the ML statistics for detecting a change in the mean of normal random variables. They found that when the initial value of the mean is unknown, which is usually the case, the Bayes-type statistic provided better power against small changes than the maximum likelihood statistics. A non-parametric approach to the change-point problem has been considered by Csörgő and Horváth (1988); see Csörgő and Horváth (1997) for an extensive review.

Interest in variance changes seems to start from the Wichern, Miller and Hsu (1976) paper, where they offer a "moving-block" procedure for detecting step changes of variance in residuals obtained from a fitted univariate AR model of order one. Baufays and Rasson (1985) improve on the Wichern et al. procedure by handling several points of variance change simultaneously. Hsu (1977, 1979) looked at detecting a single change in variance at an unknown point in time in a series of independent observations. Davis (1979) studied tests for a single change in the innovations variance at a specified point in time in a univariate AR process. Abraham and Wei (1984) used a Bayesian framework to study changes in the innovation variance of a univariate ARMA process. Srivastava (1993) found a cumulative sum of squares (CUSUM) procedure to perform better than exponentially weighted moving average procedure for detecting an increase in variance in univariate white noise series. By contrast, the change-point problem in the finance literature has dealt with CUSUM-type tests for changes in univariate conditional variances; see, e.g., Kokoszka and Leipus (2000).

Vostrikova (1981) and others have pointed out that a method for detecting a single change may be able to detect changes by binary segmentation (BS). First the entire data set is tested. If a change is detected, then the data set is split at the most likely location and the change-point procedure is applied to each new group of data; see Subsection 4.3 for more details. Inclán and Tiao (1994) adapted the BS procedure, called iterative cumulative sum of squares algorithm, to detect and locate univariate variance change-points. Their test statistic is a CUSUM of squares test for variance changes in a given sequence of independent observations taken from a univariate normal distribution. In a similar vein, Chen and Gupta (1997) used the BS procedure combined with the Schwarz information criterion (SIC).

The problem of detecting change-points with observations taken from a multivariate normal distribution has been considered by Chen and Gupta (2000). Specifically, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be a sequence of independent m-dimensional normal random vectors with parameters $(\mathbf{0}, \Sigma_1)$, $(\mathbf{0}, \Sigma_2)$, $\ldots$, $(\mathbf{0}, \Sigma_n)$, respectively. These authors are interested in testing the hypothesis

$$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_n = \Sigma_0 \quad \text{(unknown)}. \tag{1}$$

Similar to their 1997 paper, Chen and Gupta (2000) adopt the BS procedure, which implies that in each step (1) is tested versus the alternative hypothesis:

$$H_1 : \Sigma_1 = \ldots = \Sigma_{k^*} \neq \Sigma_{k^*+1} = \ldots = \Sigma_n \tag{2}$$

where $m < k^* < n - m$, k^* is the unknown position of the change-point. The proposed test statistic is a likelihood ratio (LR) test combined with either SIC or an unbiased version of it (SIC_u); see Section 3.2.

In this paper, we shall consider and apply the Bayes procedure for the detection of change-points in multidimensional stochastic processes. The resultant general test statistic will be defined in terms of sequences of partial sums of the efficient score functions of the densities from which the multivariate observations are taken. We give selected percentage points of the test statistic under the null distribution. Next, the test statistic will be specialized to the case of testing (1) versus (2).

The paper is organized as follows. In Section 2 the proposed test statistic is introduced. In Section 3 the multivariate test statistic is specialised to the case of testing changes in covariance matrices. Further, we briefly review the main features of the LR-SIC and LR-SIC_u approaches of Chen and Gupta (2000). Using simulations, we compare the finite-sample distribution and empirical power of the various test statistics in Section 4. Also, the performance of each test statistic used in conjunction with the BS procedure is investigated. After given an empirical example in Section 5, we close with a brief summary in Section 6.

2 Testing multivariate parameter constancy

2.1 Hypotheses

Suppose that $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is a finite sequence of n consecutively observed independent and identically distributed (i.i.d.) random variables taken from an m-parameter density $f(\mathbf{x}|\boldsymbol{\theta})$ where the parameter vector $\boldsymbol{\theta}$ is an m-dimensional vector such that $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)' \in \Theta \subseteq \mathbb{R}^m$, with $n > m$. The problem is to test the hypothesis

$$H_0^* : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \ldots = \boldsymbol{\theta}_n = \boldsymbol{\theta}_0 \tag{3}$$

versus the alternative

$$H_1^*: \exists\, m < k^* < n - m \text{ with } \theta_1 = \ldots = \theta_{k^*} \neq \theta_{k^*+1} = \ldots = \theta_n$$

where $\theta_0 = (\theta_0^{(1)}, \ldots, \theta_0^{(m)})'$ denotes the initial value of the parameter vector θ with $\theta_0^{(j)}$ ($j = 1, \ldots, m$) the initial value of the j-th parameter at observation point 0. It is important to distinguish between the case θ_0 is known and θ_0 is unknown.

2.2 Test statistic

Let ω_i be an indicator random variable, taking a value 1 if there is a change in θ_j between the i-th and the $(i+1)$-th observation points. The associated probabilities are defined by $P(\omega_i = 1) = p_i$ and $P(\omega_i = 0) = 1 - p_i$, with $\sum_{i=1}^{n-1} p_i = 1$. Also, let $\delta_{j,i}$ denote the amount of change in the values of the parameter vector θ_j in the interval $(i, i+1)$ ($i = 1, \ldots, n; j = 1, \ldots, m$). Assume that $\delta_{1,i}, \ldots, \delta_{m,i}$ are independent for every i, and $\delta_{j,1} \ldots, \delta_{j,n-1}$ are i.i.d. Clearly, the variables $\delta_{j,i}$, ω_i are considered nuisance parameters since their values are unknown. To derive a statistic for testing (3) we use a Bayesian method, originally due to Chernoff and Zacks (1964), that puts a priori distributions on the nuisance parameters and on the change point k^*. Then, under H_1^*, eliminate them by integration from the joint density to obtain an unconditional likelihood of $\mathbf{x}$.

Let $\delta_j = (\delta_{j,1}, \ldots, \delta_{j,n-1})$ ($j = 1, \ldots, m$), and $\omega = (\omega_1, \ldots, \omega_{n-1})$. Then, expanding the density function $f(x_i|\theta)$ under the alternative hypothesis H_1^* in a first-order Taylor expansion about θ_0 we obtain

$$\begin{aligned} f(x_i|\theta_0, \delta_1, \ldots, \delta_m, \omega) &= f(x_i|\theta_0) + \sum_{j=1} \frac{\partial}{\partial \theta_j} f(x_i|\theta_0) \Big(\sum_{u=1}^{i-1} \delta_{j,u}\omega_u \Big) + o(1) \quad (i = 1, \ldots, n) \\ &\approx f(x_i|\theta_0) \Big(1 + \sum_{j=1} \frac{\partial}{\partial \theta_j} \ln f(x_i|\theta_0) \Big(\sum_{u=1}^{i-1} \delta_{j,u}\omega_u \Big) \Big). \end{aligned}$$

Under H_0^* the joint density of $\mathbf{x}$ is given by $f_0(\mathbf{x}|\theta_0) = \prod_{i=1}^{n} f(x_i|\theta_0)$, while under H_1^* the joint density of $\mathbf{x}$ is

$$\begin{aligned} f(\mathbf{x}|\theta, \delta_1, \ldots, \delta_m, \omega) &= \prod_{i=1} f(x_i|\theta_0) \prod_{i=2} \Big(1 + \sum_{j=1} \frac{\partial}{\partial \theta_j} \ln f(x_i|\theta_0) \Big(\sum_{u=1}^{i-1} \delta_{j,u}\omega_u \Big) \Big) \\ &\approx f_0(\mathbf{x}|\theta_0) \exp \Big\{ \sum_{i=1}^{n-1} \sum_{j=1} \delta_{j,i}\omega_i \sum_{u=i+1} \frac{\partial}{\partial \theta_j} \ln f(x_u|\theta_0) \Big\}. \end{aligned}$$

We assume $\delta_j \sim N(0, \sigma_j^2)$, for some known $\sigma_j^2 > 0$, and integrate with respect to the $m(n-1)$ $\delta_{j,i}$'s ($i = 1, \ldots, n-1; j = 1, \ldots, m$) . Then, on noting that $\omega_i^2 = \omega_i$, we obtain

$$f_1(\mathbf{x}|\theta_0, \omega) \approx \frac{1}{2m} f_0(\mathbf{x}|\theta_0) \exp \Big(\frac{\sum_{j=1}^{m} \sigma_j^2}{2} \Big) \Big[1 + \sum_{i=1}^{n-1} \omega_i \sum_{j=1} \Big(\sum_{u=i+1} \frac{\partial}{\partial \theta_j} \ln f(x_u|\theta_0) \Big)^2 + o_p(\sigma_\delta^2) \Big].$$

Hence, the likelihood ratio statistic is:

$$\Lambda_n(\mathbf{x}|\omega) \equiv \frac{f_1(\mathbf{x}|\theta_0,\omega)}{f_0(\mathbf{x}|\theta_0)} = \frac{1}{2m} \exp\left(\frac{\sum_{j=1}^{m} \sigma_j^2}{2}\right)\left[1 + \sum_{i=1}^{n-1} \omega_i \sum_{j=1}^{m} \left(\sum_{u=i+1}^{n} \frac{\partial}{\partial \theta_j} \ln f(x_u|\theta_0)\right)^2\right].$$

Eliminating ω, and letting $\sigma_j^2 \to 0$ for each $j = 1, \ldots, m$, we get the two-sided test statistic

$$U_p^*(\mathbf{X}) = \sum_{\omega} P(\omega)\Lambda_n(\mathbf{x}|\omega) = \sum_{i=1}^{n-1} p_i \left(\sum_{u=i+1}^{n} v_u(\theta_0)\right)' \left(\sum_{u=i+1}^{n} v_u(\theta_0)\right), \tag{4}$$

where $v_u(\theta_0) = (\partial \ln f(x_u|\theta_0)/\partial\theta_1, \ldots, \partial \ln f(x_u|\theta_0)/\partial\theta_m)'$ is the gradient of the log-likelihood function. The hypothesis H_0^* is rejected if $U_p^*(\mathbf{X}) > c_\alpha$, where α is the given level of significance.

It may be remarked that application of the above test statistics requires that the distribution of $\mathbf{X}$ be available. One suitable family of distributions is the class of multivariate natural exponential type distributions (see, e.g., Kotz, Balakrishnan and Johnson (2000, Ch. 5)) with joint density function $f(\mathbf{x}|\theta) = h(\mathbf{x})\exp\{q_0(\theta) + \sum_{j=1}^{m} q_j(\theta)b_j(\mathbf{x})\}$, where $q_j(\theta)$ is a monotonically increasing function of θ. This family includes such distributions as the multivariate normal, discrete multivariate power series, Wishart, and bivariate Poisson.

2.3 Large-sample distribution, θ_0 is known

To obtain an asymptotic test of level α it is necessary to prove distributional convergence of (4) under H_0^*. Since under H_0^* one knows that $\{\mathbf{X}_i\}_{1\leq i\leq n}$ and $\{\mathbf{X}_{n-i+1}\}_{1\leq i\leq n}$ are equal in distribution it follows that

$$U_p^*(\mathbf{X}) \stackrel{L}{=} \sum_{i=1}^{n-1} p_{n-i} \left(\sum_{u=1}^{i} v_u(\theta_0)\right)' \left(\sum_{u=1}^{i} v_u(\theta_0)\right).$$

It turns out that the limit becomes more simple if $U_p^*(\cdot)$ is modified to

$$V_p^*(\mathbf{X}) = \sum_{i=1}^{n-1} p_{n-i} \left(\mathbf{I}^{-1/2}(\theta_0) \sum_{u=1}^{i} v_u(\theta_0)\right)' \left(\mathbf{I}^{-1/2}(\theta_0) \sum_{u=1}^{i} v_u(\theta_0)\right), \tag{5}$$

where $\mathbf{I}(\theta_0)$ is the $m \times m$ Fisher information matrix, which is assumed to exist in the neighbourhood of θ_0 in the interior of Θ.

To prove convergence in distribution we introduce a sequence of random vector functions, on the space of continuous real-valued vector functions $C_{[0,1]^m}$, as $\{\mathbf{Y}_n(t) = (Y_n^{(1)}(t), \ldots, Y_n^{(m)}(t)), t \in [0,1]\}_{n=1}^{\infty}$ possessing continuous sampling paths by the relation

$$\mathbf{Y}_n(t) = \left\{\frac{1}{\sqrt{n}} S_{n,[nt]}(\theta_0) + (nt - [nt])v_{n,[nt]+1}(\theta_0)\mathbf{I}^{-1/2}(\theta_0)\right\}, \tag{6}$$

where each component $Y_n^{(j)}(t)$ is the continuous polygonal line through the points $(\frac{i}{n}, n^{-1/2}S_{n,i}^{(j)}(\cdot))_{0\leq i\leq n}$ with $S_{n,i}^{(j)}(\cdot)$ the j-th component of

$$S_{n,i}(\theta_0) = \Gamma^{-1/2}(\theta_0)\sum_{u=1}^{i} v_u(\theta_0), \quad S_{n,0}(\theta_0) \equiv 0,$$

and with $[nt]$ the integer part of nt. Since $S_{n,i}^{(j)}(\cdot)$ are partial sums of i.i.d. random variables Donsker's Theorem states that $Y_n^{(j)}(t) \xrightarrow{\mathcal{L}} B^{(j)}(t)$ in $C_{[0,1]}$ $\forall\, 1 \leq j \leq m$, where $B^{(j)}(t)$ is the univariate standard Brownian motion. Davidson (1994, Theorem 27.17) extended this result to vector-valued processes. If mild conditions on the stochastic process $\Gamma^{-1/2}(\theta_0)v_k(\theta_0)$ are imposed, the following general result (functional central limit theorem (FCLT)) can be proved: $\mathbf{Y}_n(t) \xrightarrow{\mathcal{L}} \mathbf{B}(t) = (B^{(1)}(t), \ldots, B^{(m)}(t))$ in $C_{[0,1]^m}$ with $B^{(j)}(t)$ $(1 \leq j \leq m)$ independent.

In order to obtain the limiting distribution of the test statistic (5), assume $\{p_i\}_{i=1}^{n-1}$ to be a non-negative weight sequence on the unit interval $[0,1]$ for the unknown change-point k^* such that such that $p_{n-i} = \int_{(2i-1)/2n}^{(2i+1)/2n} \psi(t)dt$ $(i = 1, \ldots, n-1)$. If $\phi(\cdot)$ denotes the primitive of $\psi(\cdot)$ (5) can be written as

$$\begin{aligned} n^{-1}V_p^*(\mathbf{X}) &= \sum_{i=1}^{n-1}\left\{\phi\left(\frac{2i+1}{2n}\right) - \phi\left(\frac{2i-1}{2n}\right)\right\}(n^{-1/2}S_{n,i}(\theta_0))'(n^{-1/2}S_{n,i}(\theta_0)) \\ &= \sum_{i=1}^{n-1}\left\{\phi\left(\frac{2i+1}{2n}\right) - \phi\left(\frac{2i-1}{2n}\right)\right\}\mathbf{Y}_n(\frac{i}{n})'\mathbf{Y}_n(\frac{i}{n}) = h_n(\mathbf{Y}_n) \end{aligned}$$

with mapping $h_n : C_{[0,1]^m} \to \mathrm{IR}$. Thus one has to show that $\{h_n\}$ converges to h: $C_{[0,1]^m} \to \mathrm{IR}$ in the sense of Billingsley's (1968, Theorem 5.5) extended Continuous Mapping Theorem. (CMT). Here $h(\cdot)$ is given by

$$\begin{aligned} h(\mathbf{f}) &= \int_0^1 \left\{\sum_{\ell=1}^{m} f^{(\ell)}(t)\right\}^2 \phi(dt) \\ &= \int_0^1 \left\{\sum_{\ell=1}^{m} f^{(\ell)}(t)\right\}^2 \psi(t)dt = \mathrm{tr}\left\{\int_0^1 \psi(t)\mathbf{f}(t)'\mathbf{f}(t)dt\right\}, \end{aligned}$$

where $\mathbf{f}(t) = (f^{(1)}(t), \ldots, f^{(m)}(t)) \in C_{[0,1]^m}$. Then, under the assumptions of the FCLT and the extended CMT, we see that the test statistic

$$n^{-1}V_p^*(\mathbf{X}) = \sum_{i=1}^{n-1} p_{n-i}\left(\sum_{u=1}^{i} v_u(\theta_0)'\right)\Gamma^{-1}(\theta_0)\left(\sum_{u=1}^{i} v_u(\theta_0)\right) \xrightarrow{P} \mathrm{tr}\left\{\int_0^1 \psi(t)\mathbf{B}(t)\mathbf{B}'(t)dt\right\} \qquad (7)$$

where $\xrightarrow{P}$ denotes convergence in probability.

2.4 Large-sample distribution, θ_0 is unknown

In many real situations θ_0 is unknown instead of being known. Then, on replacing θ_0 in (4) by a consistent estimator, say $\hat{\theta}$, the resulting test statistic is given by

$$V_p(X) = \sum_{i=1}^{n-1} p_{n-i} \left(\sum_{u=1} v_u(\hat{\theta})'\right) \left(\sum_{u=1} v_u(\hat{\theta})\right) \tag{8}$$

where $v_u(\hat{\theta}) = (\partial \ln f(x_u|\theta)/\partial\theta_1|_{\theta_1=\hat{\theta}}, \ldots, \partial \ln f(x_u|\theta)/\partial\theta_m|_{\theta_m=\hat{\theta}})'$. Finite-sample distribution theory for (8) is complicated. Asymptotic theory is tractable. From the multivariate Taylor's expansion it follows that $\sqrt{n}(\hat{\theta} - \theta_0) = n^{-1/2} I^{-1}(\theta_0) \sum_{u=1}^{n} v_u(\theta_0) + o_p(1)$. Using the multivariate Taylor's expansion of the log-likelihood function about θ_0, we obtain $\ln f(x_u|\hat{\theta}) = \ln f(x_u|\theta_0) + \sum_{i=1}^{m} (\hat{\theta}_i - \theta_0^{(i)}) \partial \ln f(x_u|\theta_0)/\partial\theta_i + o_p(1)$. Differentiating this expression with respect to θ_i and applying Khinchine's law of large numbers, as $n \to \infty$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{u=1}^{[nt]} v_u(\hat{\theta}) &= \frac{1}{\sqrt{n}} \sum_{u=1}^{[nt]} v_u(\theta_0) - t\sqrt{n}\, I_n(\hat{\theta})(\hat{\theta} - \theta_0)' + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{u=1}^{[nt]} v_u(\theta_0) - \frac{t}{\sqrt{n}} \sum_{u=1} v_u(\theta_0) + o_p(1)
\end{aligned}$$

where $I_n(\hat{\theta}) = -n^{-1}(\partial v_u(\theta)'/\partial\theta|_{\theta=\hat{\theta}})(\partial v_u(\theta)/\partial\theta|_{\theta=\hat{\theta}})$ is the second-derivative estimate of the information matrix. This result holds uniformly in $t \in [0,1]$.

Similar as in Subsection 2.3, let $\{S_{n,u}(\hat{\theta})\}_{u=0}^{\infty}$ denote a sequence of partial sums with $S_{n,i}(\hat{\theta}) = (\sum_{u=1}^{i} v_u(\hat{\theta})) I_n^{-1/2}(\hat{\theta})$, $S_{n,0}(\hat{\theta}) \equiv 0$. Then the sequence $[\{n^{-1/2}(S_{n,u}(\hat{\theta}) - (u/n)S_{n,n}(\hat{\theta}))\}_{u=1}^{n}]_{n=1}^{\infty}$ can be used to define the sequence of random vector functions $\{\hat{Y}_n(t), t \in [0,1]\}_{n=1}^{\infty}$ as follows

$$\hat{Y}_n(t) = \frac{1}{\sqrt{n}} \left\{ S_{n,[nt]}(\hat{\theta}) - S_{n,n}(\hat{\theta})) + (nt - [nt]) v_{n,[nt]+1}(\hat{\theta}) I_n^{-1/2}(\hat{\theta}) \right\}. \tag{9}$$

In Subsection 2.3 we already showed that $Y_n(t) \xrightarrow{L} B(t) = (B^{(1)}(t), \ldots, B^{(m)}(t))$ in $C_{[0,1]^m}$. Therefore, the following FCLT holds: $\hat{Y}_n(t) \xrightarrow{L} B_0(t) = (B_0^{(1)}(t), \ldots, B_0^{(m)}(t))$ in $C_{[0,1]^m}$ where the components $B_0^{(j)}$ $(1 \leq j \leq m)$, are independent Brownian bridges. This in turn leads to $h_n(\hat{Y}_n) \xrightarrow{L} h(B_0)$ by another application of the extended CMT. Then, given (8), we see that the test statistic

$$\begin{aligned}
n^{-1} V_p(X) &= \sum_{i=1}^{n-1} p_i \left(\sum_{u=1} v_u(\hat{\theta}) - \frac{i}{n} \sum_{u=1} v_u(\hat{\theta})\right)' I_n^{-1}(\hat{\theta}) \left(\sum_{u=1} v_u(\hat{\theta}) - \frac{i}{n} \sum_{u=1} v_u(\hat{\theta})\right) \\
&\xrightarrow{D} \mathrm{tr}\left\{\int_0^1 \psi(t) B_0(t) B_0'(t)\, dt\right\}.
\end{aligned} \tag{10}$$

Table 1: Selected percentage points for $P(tr\{\int_0^1 \mathrm{B}_0(t)\mathrm{B}_0'(t)dt\} < c_\alpha)$ $(m = 1, \ldots, 5)$.

	Probabilities								
m	.01	.025	.05	.1	.5	.9	.95	.975	.99
1	.02480	.03035	.03656	.04601	.11888	.34730	.46136	.58062	.74346
2	.07883	.09362	.10941	.13222	.27757	.60704	.74752	.88799	1.07366
3	.14938	.17407	.19969	.23549	.44138	.84116	1.00018	1.16809	1.35861
4	.23104	.26555	.30066	.34862	.60668	1.06311	1.23730	1.40579	1.62263
5	.32080	.36486	.40899	.46828	.77253	1.27748	1.46466	1.64465	1.87215

The weight function $\psi(t) \equiv 1$ corresponds to a uniform prior on the unknown change-point k^*. Then from (10) we have

$$\begin{aligned} n^{-1}Q_n(m) &= (n-1)^{-1}\sum_{i=1}^{n-1}\left(\sum_{u=1}^{i} v_u(\hat{\mu}) - \frac{i}{n}\sum_{u=1}^{n} v_u(\hat{\mu})\right)' \mathrm{I}_n^{-1}(\hat{\mu})\left(\sum_{u=1}^{i} v_u(\hat{\mu}) - \frac{i}{n}\sum_{u=1}^{n} v_u(\hat{\mu})\right) \\ &\xrightarrow{D} tr\left\{\int_0^1 \mathrm{B}_0(t)\mathrm{B}_0'(t)dt\right\}, \end{aligned} \tag{11}$$

with $Q_0(m) = Q_n(m) = 0$. Test statistic $Q_n(m)$ will be the main focus of attention in the rest of the paper.

Computing quantiles for the integral in (11) can pose analytic difficulties. Anderson and Darling (1952) developed a methodology for stochastic integrals of the form in (10) with the underlying process following a one-dimensional Brownian bridge. The methodology involves identifying the sequence of eigenvalues and eigenfunctions satisfying the Fredholm integral equation. Based on these eigenvalues and the corresponding eigenfunctions, the Karhünen-Loève expansion of the univariate process $\{B_0(t), t \in [0,1]\}$ can be obtained, and the characteristic function (c.f.) of $\int_0^1 B_0^2(t)dt$ can be shown to be $\Phi(t) = \prod_{j=1}^{\infty}(1 - 2it(\pi j)^{-2})^{-1/2}$. Since $\mathrm{B}_0(t)$ is a vector Brownian bridge with independent elements, the multivariate c.f. of $tr\{\int_0^1 \mathrm{B}_0(t)\mathrm{B}_0'(t)dt\}$ is, therefore, the product of m copies of the c.f. of $\int_0^1 B_0^2(t)dt$, i.e. $\Phi_m(t) = \prod_{j=1}^{\infty}(1 - 2it(\pi j)^{-2})^{-m/2} = \left(\frac{\sin(\sqrt{2it})}{\sqrt{2it}}\right)^{-m/2}$. Given this c.f., Kiefer (1959) obtained the following expression for the limiting distribution function

$$\Delta_m(c) = \frac{2^{(m+1)/2}}{c^{m/4}(\pi)^{1/2}}\sum_{j=0}^{\infty}\frac{\Gamma(j+m/2)}{j!\,\Gamma(m/2)}e^{-(j+m/4)^2/c}D_{m-2/2}((2j+m/2)/(c^{1/2})), \quad c > 0, \tag{12}$$

where $D_a(\cdot)$ is the parabolic cylinder function. Selected quantiles are presented in Table 1 for $m = 1, \ldots, 5$.

3 Testing changes in covariance matrices

3.1 Multivariate Cramér-von Mises type test statistic

The test statistic (8) can be specialised to testing various interesting cases. For instance, assuming $\mathbf{x}$ follows a multivariate normal distribution, the test statistic can be used for testing changes in the mean vector, testing for changes in the covariance matrix, and testing for changes in the mean and covariance matrix jointly. Here we consider the case of testing for changes in covariance matrices. To allow for a comparison between our test and the test statistics given in Subsection 3.2 we assume that, under H_0, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. $N_m(0, \Sigma)$ where $\Sigma = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. Let $\hat{\sigma}_j^2 = n^{-1} \sum_{i=1}^n x_{j,i}^2$ be the ML estimator of σ_j^2. Then, it is easy to see that

$$v_u(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2) = \left(\frac{1}{-2\hat{\sigma}_1^2}\Big(1 - \frac{x_{1,u}^2}{\hat{\sigma}_1^2}\Big), \ldots, \frac{1}{-2\hat{\sigma}_m^2}\Big(1 - \frac{x_{m,u}^2}{\hat{\sigma}_m^2}\Big) \right)' \tag{13}$$

while

$$\mathbf{I}_n^{-1}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_m^2) = \mathrm{diag}\left(\frac{2\hat{\sigma}_1^4}{n}, \ldots, \frac{2\hat{\sigma}_m^4}{n} \right). \tag{14}$$

Substituting (13) and (14) into (11) the derived test statistic is:

$$Q_n(m) = \frac{n^2}{2(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{m} \left(\frac{C_{j,i}}{C_{j,n}} - \frac{i}{n} \right)^2 \tag{15}$$

where $C_{j,i} = \sum_{u=1}^{i} x_{j,u}^2 \quad (i = 1, \ldots, n; j = 1, \ldots, m)$.

When, using (15), the existence of a change-point is determined, the next problem to be considered is the estimation of the location. Let $SS_k = |\sum_{i=1}^{k} (\mathbf{x}_i - \overline{\mathbf{x}}_k)(\mathbf{x}_i - \overline{\mathbf{x}}_k)'|$ and $SS_{n-k} = |\sum_{i=1}^{n-k} (\mathbf{x}_i - \overline{\mathbf{x}}_{n-k})(\mathbf{x}_i - \overline{\mathbf{x}}_{n-k})'|$ where $\overline{\mathbf{x}}_k = k^{-1} \sum_{i=1}^{k} \mathbf{x}_i$ and $\overline{\mathbf{x}}_{n-k} = (n-k)^{-1} \sum_{i=1}^{n-k} \mathbf{x}_i$. Then an estimate of the location of the change-point, is obtained by minimizing $SS_{n,k} = SS_k + SS_{n-k}$, i.e. $SS_{k^*} = \min_{m<k<n-m} SS_{n,k}$. Generalization of this estimation procedure to the case of more than one change-point is direct.

3.2 Likelihood ratio based test statistic

Most attention in the literature has focussed on the likelihood ratio (LR) test as a way of checking H_0 against H_1. If the change at point k^* is known, then H_0 will be rejected for small values of the test statistic $\Lambda_n(\mathbf{x}) = \sup_{\theta \in \Theta} \prod_{1 \leq i \leq n} f(\mathbf{x}_i|\theta) / \sup_{\theta \in \Theta} \prod_{1 \leq i \leq k} f(\mathbf{x}_i|\theta) \prod_{k<i \leq n} f(\mathbf{x}_i|\theta)$. If the densities, as smooth functions of the parameters, can be estimated consistently it is well-known that the log likelihood ratio can be written as $-2\ln(\Lambda_n(m)) = 2\{L_k(\hat{\theta}_k) + L_{n-k}(\hat{\theta}_{n-k}) - L_n(\hat{\theta}_n)\}$,

where $L_k(\mathbf{y}) = \sum_{1 \leq i \leq k} \ln f(\mathbf{x}_i|\mathbf{y})$, and $L_{n-k}(\mathbf{y}) = \sum_{k<i \leq n} \ln f(\mathbf{x}_i|\mathbf{y})$. Since k^* is unknown, H_0 will be rejected for large values of the test statistic

$$\lambda_n(m) = \max_{2<k<n-m} \{-2\ln(\Lambda_n(m)\}. \tag{16}$$

Chen and Gupta (1995, 2000) show that the asymptotic null distribution of (16), without imposing the normality assumption, is Gumbel. Indeed, let

$$a\{\ln(n)\} = (2\ln\ln(n))^{1/2} \quad \text{and} \quad b_n\{\ln(n)\} = 2\ln\ln(n) + (m/2)\ln\ln\ln(n) - \ln\Gamma(m/2).$$

Suppose all the necessary regularity conditions hold. Then, under H_0, for all $x \in \mathbb{R}$, m fixed, $m < k < n - m$ and $n \to \infty$, $k \to \infty$, such that $k/n \to 0$, the asymptotic distribution is given by

$$\lim_{n \to \infty} P[a\{\ln(n)\}\lambda_n(m) - b_m\{\ln(n)\} \leq x] = \exp(-2e^{-x}).$$

An approximate expression for the critical value $c_\alpha(n,m)$ is given by

$$\begin{aligned} c_\alpha(n,m) &= \left\{ -\frac{1}{a\{\ln(n)\}} \ln\ln\left[1 - \alpha + \exp\{-2\exp(b_m\{\ln(n)\})\}\right]^{-1/2} + \frac{b_m\{\ln(n)\}}{a\{\ln(n)\}} \right\}^2 \\ &\quad - \frac{1}{2}m(m+1)\ln(n) \end{aligned} \tag{17}$$

with α the Type I error. Chen and Gupta (1994, 2000) provide tables with values of (17) for selected values of α, n, and m. Also, as a special case of the LR test procedure, these authors propose the following test statistic for testing H_0 versus H_1:

$$\lambda_n(m) = \max_{m<k<n-m} (n\ln|\hat{\Sigma}_n| - k\ln|\hat{\Sigma}_k| - (n-k)\ln|\hat{\Sigma}_{n-k}|)^{1/2} \tag{18}$$

where $\hat{\Sigma}_n = n^{-1}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'$, $\hat{\Sigma}_k = k^{-1}\sum_{i=1}^{k} \mathbf{x}_i\mathbf{x}_i'$, and $\hat{\Sigma}_{n-k} = (n-k)^{-1}\sum_{i=k+1}^{n} \mathbf{x}_i\mathbf{x}_i'$ are the ML estimators of Σ_n, Σ_k and Σ_{n-k} respectively. Note that to be able to obtain these ML estimators, the testing procedure can detect only changes for $m < k < n - m$.

It is easy to see that

$$\lambda_n^2(m) + \frac{1}{2}m(m+1)\ln(n) = -\max_{m<k<n-m} \{\mathrm{SIC}(k,m) - \mathrm{SIC}(n,m)\}$$

where $\mathrm{SIC}(n,m)$, under H_0, is given by

$$\mathrm{SIC}(n,m) = mn\ln(2\pi) + n\ln|\hat{\Sigma}_n| + n + \frac{m}{2}(m+1)\ln(n)$$

and where $\mathrm{SIC}(k,m)$, under H_1, is

$$\mathrm{SIC}(k,m) = mn\ln(2\pi) + k\ln|\hat{\Sigma}_k| + (n-k)\ln|\hat{\Sigma}_{n-k}| + n + m(m+1)\ln(n).$$

Then, on adopting the binary segmentation method, H_0 is not rejected if, for some k,

$$\mathrm{SIC}^*(n,m) \equiv \mathrm{SIC}(n,m) - \min_{m<k<n-m} \mathrm{SIC}(k,m) \leq c_\alpha(n,m), \qquad (c_\alpha(n,m) > 0), \tag{19}$$

otherwise reject H_0.

As an alternative to the SICs, Chen and Gupta (2000) derive two approximate unbiased estimators of the Kullback-Leibler information. Under H_0 and H_1 these estimators are respectively given by

$$\mathrm{SIC}_u(n,m) = \mathrm{SIC}(n,m) + \frac{n(mn - n + m + 1)}{n - m - 1} - \frac{1}{2}m(m+1)\ln(n),$$
$$\mathrm{SIC}_u(k,m) = \mathrm{SIC}(k,m) + \frac{k^2 m(n - k - m - 1) + (n-k)^2 m(k - m - 1)}{(k - m - 1)(n - k - m - 1)} - n - m(m+1)\ln(n).$$

In this case H_0 is not rejected if, for some k,

$$\mathrm{SIC}_u^*(n,m) \equiv \mathrm{SIC}_u(n,m) - \min_{m<k<n-m} \mathrm{SIC}_u(k,m) \leq c_\alpha(n,m), \qquad (c_\alpha(n,m) > 0), \tag{20}$$

otherwise reject H_0.

Table 2: Null distribution of $Q_n(m)$ using quantiles of Table 1; $\mathrm{SIC}^*(n,m)$; and $\mathrm{SIC}_u^*(n,m)$; $\{X_i\}_{i=1}^n \sim$ i.i.d. $N(0, I_m)$ with m = 1, 2, 3, and n = 25, 50, and 100; 50,000 replications.

Exact	Test	m = 1			m = 2			m = 3		
Prob.	Statistic	n = 25	n = 50	n = 100	n = 25	n = 50	n = 100	n = 25	n = 50	n = 100
.900	$Q_n(m)$	.906	.902	.899	.916	.907	.903	.923	.910	.904
	$\mathrm{SIC}^*(n,m)$	.884	.885	.884	.955	.974	.985	.980	.993	.998
	$\mathrm{SIC}_u^*(n,m)$	.904	.838	.754	.983	.954	.918	.999	.991	.977
.950	$Q_n(m)$	.953	.952	.951	.960	.955	.951	.965	.956	.953
	$\mathrm{SIC}^*(n,m)$	.941	.943	.943	.975	.986	.992	.988	.996	.999
	$\mathrm{SIC}_u^*(n,m)$	.952	.920	.873	.991	.976	.955	.999	.996	.987
.975	$Q_n(m)$	.977	.977	.976	.982	.979	.976	.984	.979	.978
	$\mathrm{SIC}^*(n,m)$	.971	.972	.971	.986	.992	.996	.993	.998	.999
	$\mathrm{SIC}_u^*(n,m)$	.976	.960	.937	.995	.987	.977	1.000	.998	.993
.990	$Q_n(m)$	.992	.991	.990	.994	.992	.990	.997	.995	.994
	$\mathrm{SIC}^*(n,m)$	.988	.989	.988	.993	.996	.998	.996	.999	.999
	$\mathrm{SIC}_u^*(n,m)$	.990	.984	.974	.998	.995	.990	1.000	.999	.997

4 Simulation results

The performance of the various multidimensional test statistics for detecting changes in covariance matrices will be evaluated in several ways. First, we compare the null distributions of

$Q_n(m)$, SIC$^*(n, m)$, and SIC$^*_u(n, m)$. Next, we evaluate the empirical power of these test statistics for detecting changes in covariance matrices. Finally, we present results on the performance of the test statistics jointly with the use of the BS procedure to determine the "correct" number of changes.

4.1 Null distribution

Table 2 compares the finite-sample null distributions of the test statistics $Q_n(m)$, SIC$^*(n, m)$, and SIC$^*_u(n, m)$ with their asymptotic distributions. Only results for the upper tails are reported, which are more relevant from a hypothesis testing viewpoint. We can conclude that the finite-sample distribution of $Q_n(m)$ can be reasonably accurately approximated by the asymptotic distribution (12) for all values of n, and irrespective of the value of m. On the other hand, for the test statistics SIC$^*(n, m)$ and SIC$^*_u(n, m)$ the upper tails are too large. As a result, we expect a problem of overrejection. Nevertheless, some improvement in the performance of these test statistics can be noticed when n increases, for fixed values of m.

4.2 Empirical power

We now examine the power of the three test statistics under the alternative hypothesis

$$\begin{aligned} H_1: \ \{X_i\} &\sim \ i.i.d.\, N(0, I_m), \quad (i = 1, \ldots, [n/2]), \\ \{X_i\} &\sim \ i.i.d.\, N(0, \sigma I_m), \quad (i = [n/2], \ldots, n), \end{aligned}$$

where $[n/2]$ is the point where the change in the covariance matrix occurs, and $\sigma > 0$ denotes a fixed scale factor. As a start, using 50,000 replications of $\{X_i\}_{i=1}^n \sim i.i.d.\, N(0, I_m)$, we empirically calculated the cut point for each test statistic and each sample size having the common level of significance $\alpha = .05$. Next, from 50,000 replications of the process under H_1 we determined the size-adjusted power of the tests, the size being common for all four test statistics.

The graphs in Figure 1 show power functions for $n = 50$ and $n = 100$, with $\sigma = 1, 1.5, 2, \ldots, 4$. It is clear that the test statistic $Q_n(m)$ is more powerful than SIC$^*(n, m)$ and SIC$^*_u(n, m)$ for low values of σ. Unsurprisingly, the powers of all test statistics increase as n increases. However, it is interesting to see that the powers of $Q_n(m)$ strongly depends on the value of m. The gain in power seems to be more due to an increase in the dimension m than to an increase in the sample size n. Further, as expected, the test statistic SIC$^*_u(n, m)$ has higher power than SIC$^*(n, m)$.

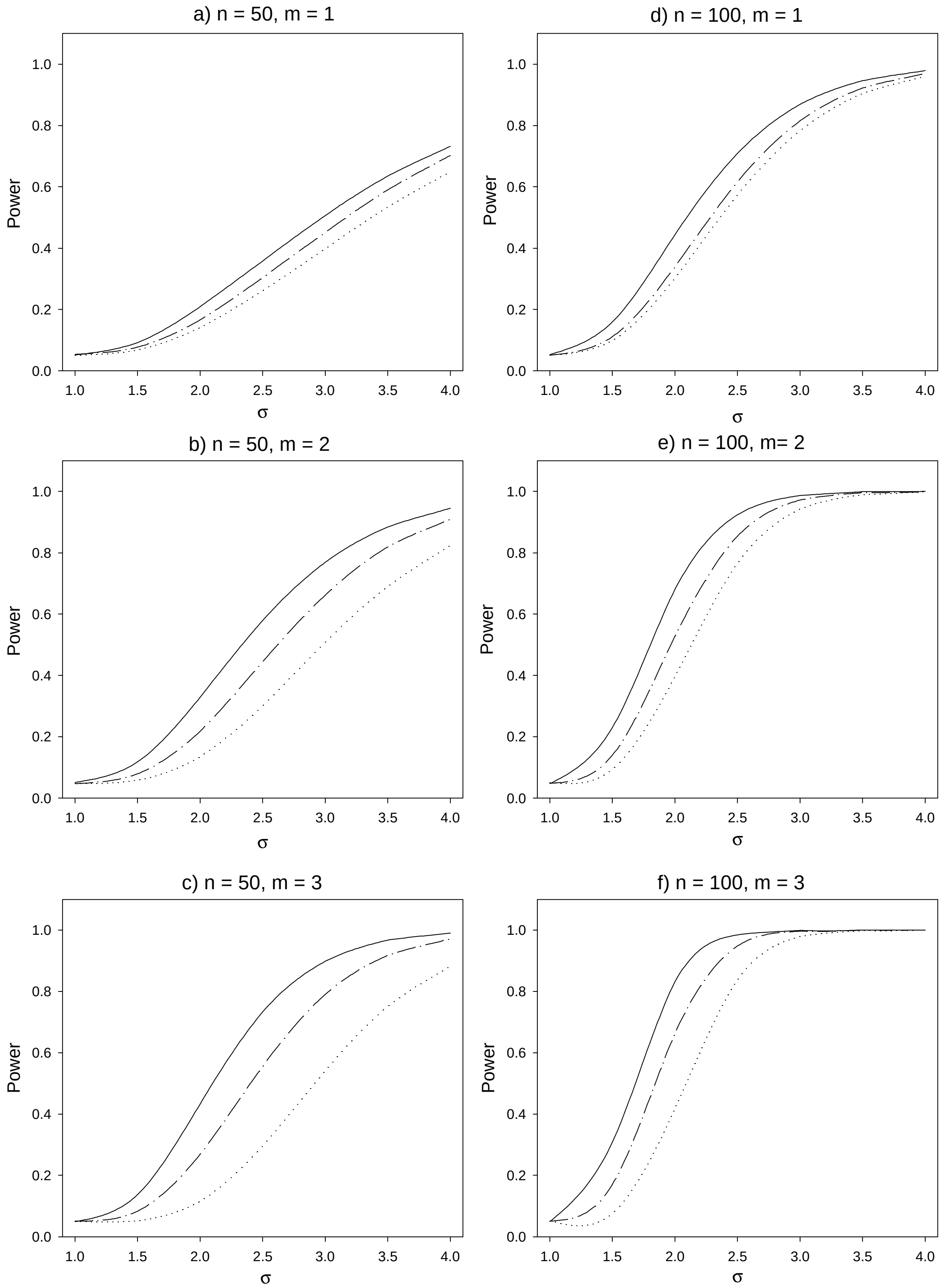

Figure 1: Power functions for the test statistics $Q_n(m)$ (solid line), SIC$^{¤}(n, m)$ (dotted line), and SIC$^{¤}_{u}(n, m)$ (dash-dotted line); i.i.d. case.

4.3 Binary segmentation procedure

The BS procedure, as proposed by Vostrikova (1981), proceeds as follows in an m-dimensional setting:

1. Begin by calculating the value of a change-point statistic, say $T(t_1, n, m)$, where t_1 is the start point and n is the endpoint of the series. Suppose we observe a statistically significant value at the point k_1.

2. Search using the statistic $T(t_1, k_1, m)$, looking for a significant value in the first half of the series determined by the partition. If a statistically significant value is found, say k_2, repeat this step using $T(t_1, k_2, m)$. Continue until no more statistically significant values are found in the interval formed each time by the range from the beginning of the series to the current maximum.

3. Repeat the procedure by beginning at the first point found and working to the end of the series. For example, let ℓ_1 be the point that gives a significant value of $T(k_1, n, m)$ and search for a significant value of $T(\ell_1, n, m)$. Call the point ℓ_2.

To compensate for an apparent overestimation of the number of change-points Inclán and Tiao (1994) included the following additional step to the above procedure:

4. Arrange the change-points found in order, say $k_1^* < k_2^* < \ldots < k_{\text{last}}^* < \ell_1^* < \ell_2^* < \ldots < \ell_{\text{last}}^*$ and check the location of each point by straddling with the test statistic $T(\cdot)$. For example, check k_2 using $T(k_1, k_3, m)$. Repeat this step until the number of change-points does not change and the points found in each new pass are "close" to those in the previous pass.

Simulations were run with this additional step, assuming that the BS procedure has converged if each change-point is within three observations of where it was on the previous iteration. Critical values of the test statistics $Q_n(m)$ are taken from Table 1. Expression (17) was used to compute critical values of the test statistics $\text{SIC}^*(n, m)$ and $\text{SIC}_u^*(n, m)$.

To assess the performance of the BS procedure we consider testing $H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_n = \mathbf{I}_m$ against $H_1 : \Sigma_1 = \ldots = \Sigma_{k^*} = \mathbf{I}_m, \Sigma_{k^*+1} = \ldots = \Sigma_n = \sigma \mathbf{I}_m$ with $k^* = [n/4]$, $[n/2]$, and $[3n/4]$, σ = 1, 1.5 and 2, n = 50 and 100, and m = 1, 2, and 3. Table 3 gives the frequency distribution of the number of change-points detected by the three test statistics at location $k^* = [n/2]$. Columns corresponding to the "correct identifications" are set in bold type.

We see that for the test statistic $Q_n(m)$ the BS procedure does quite well at locating the single change-point when n = 50 and σ = 1.5. This is still the case with m = 2. However, with

Table 3: Frequency distribution of the number of detected change-points, obtained with the BS procedure, at location $k^{*} = [n/2]$; 10,000 replications, 5% critical values of the test statistics.

		Test	m = 1			m = 2			m = 3		
σ	n	Statistic	0	1	> 2	0	1	> 2	0	1	> 2
1	50	$Q_n(m)$	.957	.043	.000	.976	.024	.000	.976	.024	.000
		SIC$^{*}(n,m)$	.995	.005	.000	.988	.011	.001	.983	.016	.001
		SIC$^{*}_u(n,m)$	.994	.005	.001	.961	.035	.004	.902	.081	.017
1.5		$Q_n(m)$	.087	.907	.006	.019	.978	.003	.395	.601	.004
		SIC$^{*}(n,m)$	.347	.652	.001	.050	.942	.008	.789	.203	.008
		SIC$^{*}_u(n,m)$	.295	.704	.001	.013	.792	.195	.283	.584	.133
2		$Q_n(m)$	.001	.987	.012	.000	.996	.004	.002	.979	.019
		SIC$^{*}(n,m)$	.002	.996	.002	.000	.977	.023	.004	.943	.053
		SIC$^{*}_u(n,m)$	.001	.998	.001	.000	.663	.337	.001	.710	.289
1	100	$Q_n(m)$	.953	.047	.000	.977	.023	.000	.972	.028	.000
		SIC$^{*}(n,m)$	.996	.004	.000	.984	.015	.001	.991	.009	.000
		SIC$^{*}_u(n,m)$	.985	.015	.000	.871	.120	.009	.763	.201	.036
1.5		$Q_n(m)$	.003	.991	.006	.001	.995	.004	.014	.976	.010
		SIC$^{*}(n,m)$	.016	.980	.004	.000	.989	.011	.348	.635	.017
		SIC$^{*}_u(n,m)$	.007	.988	.005	.001	.900	.099	.001	.905	.094
2		$Q_n(m)$	.000	.988	.012	.001	.994	.005	.001	.983	.016
		SIC$^{*}(n,m)$	.000	.996	.004	.000	.984	.016	.000	.970	.030
		SIC$^{*}_u(n,m)$	.000	.994	.006	.000	.872	.128	.000	.887	.113

Table 4: Summaries from the sampling distribution of $k^{¤}$ for n = 100; 10,000 replications.

	Test	m = 1			m = 2			m = 3		
σ	Statistic	Mode (Freq.)	Mean	SD	Mode (Freq.)	Mean	SD	Mode (Freq.)	Mean	SD
1.5	$Q_n(m)$	50 (2779)	50.94	7.69	50 (4418)	50.43	4.17	50 (2303)	50.34	24.25
	SIC$^{¤}(n, m)$	50 (2937)	51.01	8.18	50 (4576)	48.29	3.74	50 (2153)	65.18	26.15
	SIC$^{¤}_u(n, m)$	50 (2873)	50.54	6.69	50 (4138)	45.00	11.71	50 (3396)	44.87	10.26
2	$Q_n(m)$	50 (5178)	50.10	4.96	50 (7424)	50.00	3.22	50 (6227)	49.33	6.55
	SIC$^{¤}(n, m)$	50 (5436)	49.50	2.08	50 (7385)	47.80	3.67	50 (6354)	46.40	4.38
	SIC$^{¤}_u(n, m)$	50 (5326)	49.45	2.42	50 (6507)	43.39	13.00	50 (5838)	44.16	9.42

m = 3, σ = 1.5 and test statistic $Q_n(m)$, the BS procedure errs towards zero change-points. For σ = 2 the BS procedure gives too many correct identi…cations, irrespective of value of the sample size n. The performance of the BS procedure improves notably for n = 100. With m = 3, n = 50 and σ = 1.5, the BS procedure errs on the side of two or more change-points with considerable frequency. This phenomenon is still present for increasing values of n and σ. Clearly, the test statistics SIC$^{¤}(n, m)$ and SIC$^{¤}_u(n, m)$ produce very few correct identi…cations when both n and σ are small. Some improvements occur for series of 100 observations, particularly for SIC$^{¤}_u(n, m)$.

Our simulation experiment also provides information about the sampling distribution of $k^{¤}$. To safe space, we summarize the main results in Table 4. Again we see the poor performance of SIC$^{¤}(n, m)$ and SIC$^{¤}_u(n, m)$, con…rming earlier results presented in Table 3. On the other hand, the test statistic $Q_n(m)$ has a sampling distribution of $k^{¤}$ with both modal and mean values exactly at the change-point 50. When the location $k^{¤}$ lays near the beginning or the end of the series, say at $k^{¤}$ = [n/4] and $k^{¤}$ = [3n/4], it is much harder to detect a single change-point than when the change-points are in the middle of the series. Nevertheless, when either σ or the sample size n increases, the number of correctly identi…ed changes via the BS procedure increases rapidly. This applies to all three test statistics.

5 Example

The Danish …re data is a set of 2493 trivariate observations consisting in losses to buildings (X_1), losses to contents (X_2), and losses to pro…ts (X_3), expressed in millions of Danish Kroner (1985 prices) covering the period 1980-1993. The data are available from www.math.eth.ch/%7EMneil/

data.html. Embrechts, Klüppelberg and Mikosch (1997, Example 6.2.9), among others, analysed the data in a one-dimensional setting. In our analysis we only use observations with strictly positive components, resulting in $n = 517$ observations. Using a homogeneous (i.e. no-change) Gumbel copula, Blum, Dias and Embrechts (2002) obtained a good fit to various functionals of the two shortened series X_2 and X_3.

Figure 2 shows a plot of the series. Also exhibited are the estimated locations of variance ($m = 1$) change-points using the three test statistics and the BS procedure jointly (5% critical values). Table 5 presents similar results for all combinations of the variables x_1, x_2, and x_3. The bursts of increased variance in all three series near location point 8 is observed by $Q_n(m)$ and when the dimensionality of the tests is set at $m = 2$ and $m = 3$. Further we see from Table 5 that the test statistics $\mathrm{SIC}^*(n,m)$ and $\mathrm{SIC}_u^*(n,m)$ quite often pick the large burst at location point 499 in series x_2 as a single change-point. Overall, the BS procedure in conjunction with the three test statistics does a good job of isolating the two main locations of variance changes in both the two and three-dimensional system.

All above results hint at possible heterogeneity of variance in the set of variables. This seems to invalidate the approach taken in Blum et al. (2002). Interestingly, when we analysed the log-transformed variables most test statistics indicated zero variance change-points; see Table 5. Exceptions were noted for the test statistic $Q_n(m)$ and the test statistic $\mathrm{SIC}_u^*(n,m)$. Thus, the assumption of homogeneity of variance may be better satisfied using the log-transformed data. Note that in theory, taking a log transformation or no log transformation is not an issue since a copula is invariant under continuous strictly increasing transformations like the log. In practice, however, the results can lead to different locations of change-points as has been shown above.

6 Summary

We have presented a multivariate Cramér-von Mises type test statistic for detecting change-points in general multidimensional stochastic processes. We have demonstrated by simulation that the test statistic has a satisfactory null distribution as compared to two other tests for change-point detection. The power characteristics are also satisfactory. On the other hand, our simulation results indicate that the LR-SIC and LR-SIC_u test statistics for locating inhomogeneities in covariance matrices perform quite poorly under various alternatives. Using a binary segmentation procedure, we have compared the various test statistics using simulated and real data. The multivariate Cramér-von Mises type test statistic performs quite well at locating

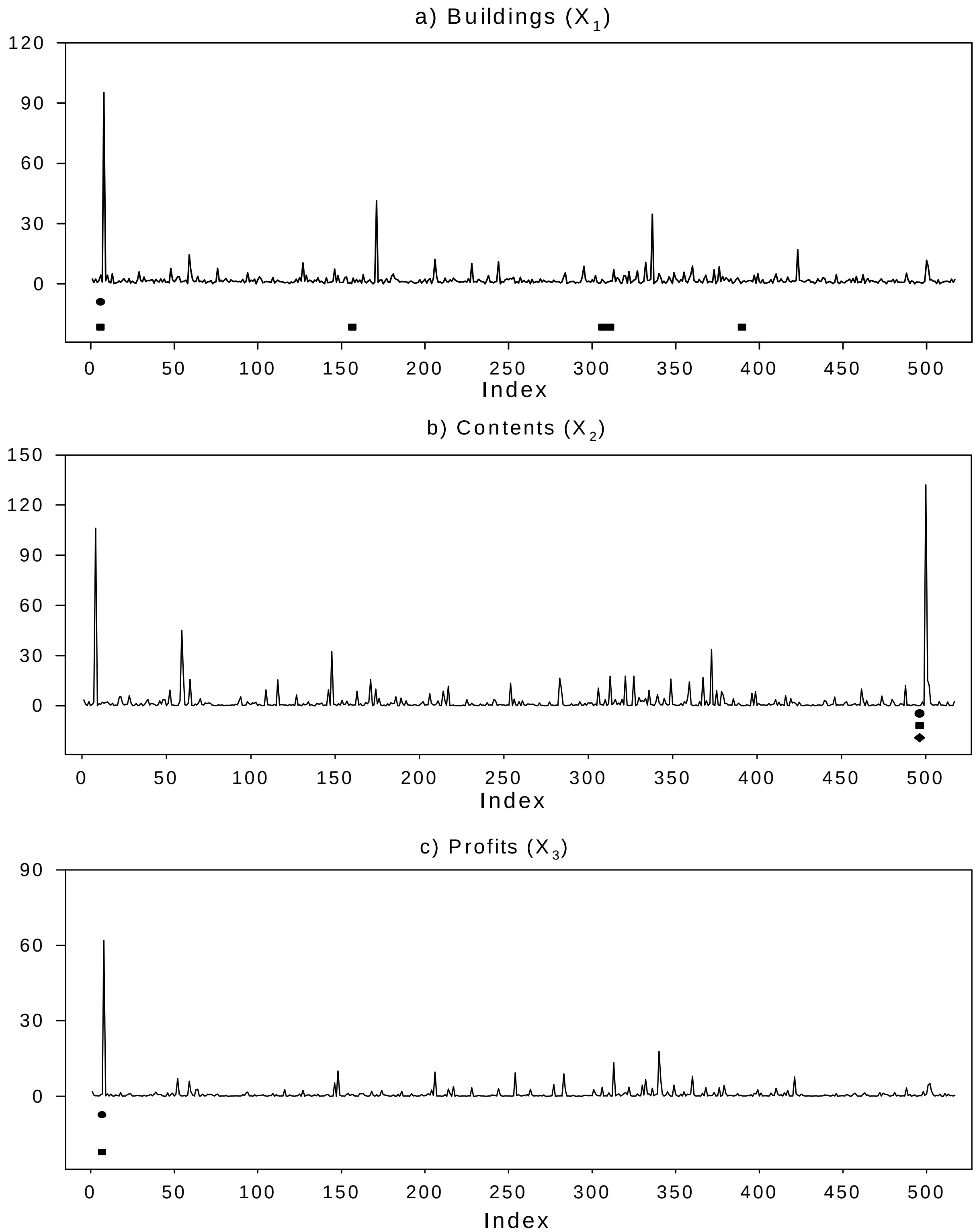

Figure 2: Danish …re data with estimated locations of variance (m = 1) change-points for each test using the BS procedure; $Q_n(m)$ (dots); $SIC^{¤}(n,m)$ (squares); and $SIC^{¤}_u(n,m)$ (diamonds).

Table 5: Locations of change-points in covariance matrices for the Danish …re data; 5% critical values of test statistics.

Test Statistic	Variables			
	(X_1, X_2) (ln X_1, ln X_2)	(X_1, X_3) (ln X_1, ln X_3)	(X_2, X_3) (ln X_2, ln X_3)	(X_1, X_2, X_3) (ln X_1, ln X_2, ln X_3)
$Q_n(m)$	8	8	8	8
	(10)	(10)	(352)	(10)
SIC$^*(n, m)$	499	7	8	499
SIC$_u^*(n, m)$	499		499	10, 499
			(312)	

change-points. Note that the normality assumption introduced in Section 3 for deriving the test statistic $Q_n(m)$ is not necessarily required for obtaining its limiting distribution. Further, the limiting distributions of (15) as derived by Kiefer (1959) makes use of Donsker's invariance principle. Since the invariance principle holds in weakly dependent processes (cf. Billingsley, 1968) it would be interesting to extend the test statistic $Q_n(m)$ to autocorrelated observations.

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.

Anderson, T.W. and Darling, D.A. (1952), "Asymptotic theory of certain 'goodness of …t' criteria based on stochastic processes", Ann. Math. Statist. , 23, 193-212.

Andreou, E. and Ghysels, E. (2002), "Detecting multiple breaks in …nancial market volatility dynamics", J. Applied Econometrics, 17, 579-600.

Baufays, P. and Rasson, J.P. (1985), "Variance changes in autoregressive models". In Time Series Analysis: Theory and Practice 7, ed. O.D. Anderson, North-Holland: New York, pp. 119-127.

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

Blum, P. , Dias, A., and Embrechts, P. (2002), "The ART of dependence modelling: the latest advances in correlation analysis". In Alternative Risk Strategies (M. Lane, ed.). Risk Books: London, pp. 339-356.

Chen, J. and Gupta, A.K. (1997), "Testing and locating variance changepoints with application to stock prices”, J. Amer. Statist. Ass., 92, 739-747.

Chen, J. and Gupta, A.K. (1995), “Change point analysis of mean vector and covariance matrix under normality”. Department of Mathematics and Statistics, Bowling Green State University, Technical Report No. 95-02.

Chen, J. and Gupta, A.K. (2000). Parametric Statistical Change Point Analysis, Birkhäuser: Boston.

Chernoff, H. and Zacks, S. (1964), “Estimating the current mean of a normal distribution which is subject to changes in time”, Ann. Math. Statist., 35, 999-1018.

Csörgő, M. and Horváth, L. (1988) Non-parametric methods for change-point problems. In Handbook of Statistics, Vol. 7 (P.R. Krishnaiah and C.R. Rao, eds.), pp. 403-425, New York: North-Holland.

Csörgő, M. and Horváth, L. (1997). Limit Theorems in Change-Point Analysis, Wiley: Chicester,

Davidson, J. (1994). Stochastic Limit Theory, Oxford University Press: Oxford.

Davis, W.W. (1979), “Robust methods for detection of shifts of the innovation variance of a time series”, Technometrics, 21, 313-320.

Embrechts, P. Klüppelberg, C., and Mikosch, T. (1997). Modelling Extremal Events for Insurance and Finance, Springer: Berlin.

Esterby, S.R. and El-Shaarawi, A.H. (1981), “Inference about the point of change in a regression model”, J. Appl. Statist., 30, 277-285.

Gardner, L.A. (1969), “On detecting changes in the mean of normal variates”, Ann. Math. Statist., 72, 180-186.

Hawkins, D.M. (1977), “Testing a sequence of observations for a shift in location”, J. Amer. Statist. Ass., 72, 180-186.

Hinkley, D.V. (1971), “Inference about the change point from cumulative sums tests”, Biometrika, 58, 509-523.

Horváth, L. (1993), “Maximum likelihood method for testing changes in the parameters of normal observations”, Ann. Statist., 21, 671-680.

Hsu, D.A. (1977), “Test for variance shifts at an unknown time point”, Appl. Statistics, 26, 179-184.

Hsu, D.A. (1979), “Detecting shifts in gamma sequences; with application to stock price and air traffic flow analysis”, J. Amer. Statist. Ass., 74, 31-40.

Inclán, C. and Tiao, G.C. (1994), “Use of cumulative sums of squares for retrospective detection of changes of variance", J. Amer. Statist. Ass., 89, 913-923.

Kander, Z. and Zacks, S. (1966), "Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points", Ann. Math. Statist., 37, 1196-1210.

Kiefer, J. (1959), "K-sample analogues of the Kolmogorov-Smirnov and Cramer V. Mises Tests", Ann. Math. Statist., 30, 420-447.

Kokoszka, P. and Leipus, R. (2000), "Change-point estimation in ARCH models", Bernouilli, 6, 1-28.

Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000). Continuous Multivariate Distributions Vol. 1: Models and Applications (2nd ed.), Wiley, New York.

Lee, S. and Wei, C-Z. (1999), "On residual empirical processes of stochastic regression models with applications to time series", Ann. Statist., 27, 237-261.

McNeil, A.J. (1967), "Estimating the tails of loss severity distributions using extreme value theory", Astin Bulletin, 27, 117-137.

MacNeill, I.B. (1978), "Properties of partial sums of polynomial regression residuals with application to tests for change in regression at unknown times", Ann. Statist., 6, 422-433.

Milbrodt, H. and Strasser, H. (1990), "On the asymptotic power of the two-side Kolmogorov-Smirnov test", J. Stat. Planning Inf., 26, 1-23

Quandt, R.E. (1958), "The estimation of the parameters of a linear regression system obeying two separate regimes", J. Amer. Statist. Assoc., 53, 873-880.

Quandt, R.E. (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", J. Amer. Statist. Assoc., 55, 324-330.

Sen, A.K. and Srivastava, M.S. (1973), "On multivariate tests for detecting change in the mean", Sankyā, A35, 173-185.

Sen, A.K. and Srivastava, M.S. (1975), "On tests for detecting change in mean", Ann. Statist., 3, 98-108.

Srivastava, M.S. (1993), "Comparison of CUSUM and EWMA procedures for detecting a shift in the mean or an increase in the variance", J. Appl. Statist. Science, 1, 445-468.

Vostrikova, L.J. (1981), "Detecting "disorder" in multidimensional random processes", Soviet Mathematics Doklady, 24, 55-59.

Wichern, D.W., Miller, R.B. and Hsu, D.A. (1976), "Changes of variance in …rst order autoregressive time series models - with an application", Appl. Statistics, 25, 248-256.

Worsley, K.J. (1979), "On the likelihood ratio tests for a shift in location of normal population", J. Amer. Statist. Ass., 74, 365-367.