A Change Point Approach for Phase I Analysis in Multistage Processes

Changliang ZOU

LPMC and Department of Statistics
School of Mathematical Sciences
Nankai University
Tianjin, China

Fugee TSUNG

Department of Industrial Engineering and Logistics Management
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
(*season@ust.uk*)

Yukun LIU

LPMC and Department of Statistics
School of Mathematical Sciences
Nankai University
Tianjin, China

We study the phase I analysis of a multistage process where the input of the current process stage may be closely related to the output(s) of the earlier stage(s). We frame univariate observations from each of the stages in a multistage process as a single vector and recognize that the directions in which these vectors can shift are limited when attention is restricted to a single step shift in the mean of one stage. This allows us to focus detection power on a limited subspace with improved sensitivity. Taking advantage of this particular characteristic, we propose a change point approach that integrates the classical binary segmentation test with the directional information based on the state-space model for testing the stability of a batch of historical data. We give an accurate approximation for the significance level of the proposed test. Our simulation results show that the proposed approach consistently outperforms existing methods for multistage processes.

KEY WORDS: Binary segmentation; Directional information; Generalized likelihood ratio test; Statistical process control.

1. INTRODUCTION

Many manufacturing operations involve multiple operation stages instead of just a single stage. For instance, in semiconductor manufacturing, Kim and May (1999) investigated the via formation procedure to produce multichip module dielectric layers composed of photosensitive benzocyclobutene. The via formation process involves several sequential stages, such as spin coating, soft baking, exposing, developing, curing, and plasma descumming. In such a case, outputs from operations at downstream stages can be affected by operations at upstream stages. In addition, a product part that is transferred from one stage to the next in a multistage process may introduce extra variations that do not occur in a generic single-stage process. Shi (2006) has provided an extensive review on the multistage process control problem with many industrial examples.

Thanks to recent advances in sensor and information technologies, automatic data acquisition techniques are commonly used in increasingly complicated processes with multiple stages. In addition, a large amount of data and information related to quality measurements in a multistage process has become available. Thus engineering and statistical approaches to make use of the multistage data and information regarding control and monitoring have become possible and beneficial in industrial practice.

One way to use the multistage information and capture the multistage relationships (e.g., the correlation structure between serially dependent stages) is to model the multistage process as a linear state-space model. Examples have been given by Jin and Shi (1999), Ding, Shi, and Ceglarek (2002b), Ding, Ceglarek, and Shi (2002a), Ding, Jin, Ceglarek, and Shi (2005), Djurdjanovic and Ni (2001), Huang, Zhou, and Shi (2002), Zhou, Ding, Chen, and Shi (2003a), and Zhou, Huang, and Shi (2003b). Even though the state-space modeling approach is popular and has been particularly influential in recent years, investigations of multistage statistical process control (SPC) based on such a model, including both phase I retrospective analysis and phase II monitoring and diagnosis, are rare. Xiang and Tsung (2006) and Zantek, Gordon, and Robert (2006) considered online monitoring and diagnosis of out-of-control (OC) conditions in the multistage process. Zhou, Chen, and Shi (2004) estimated the process parameters by applying the typical estimation methods of general mixed linear models. Xiang and Tsung (2006) alternatively derived an EM procedure for maximum likelihood (ML) estimates of the parameters.

The estimation methods proposed by Zhou et al. (2004) and Xiang and Tsung (2006) rely on the assumption that the reference sample is statistically stable. If the collected samples are indeed not in control (IC) (i.e., change points or outliers exist), then the estimates of the parameters will be adversely affected. Therefore, a phase I analysis of a multistage process, aimed at identifying the data from an IC process as accurately as possible so that quality engineers can have a good reference for establishing control charts for future operations, is very valuable and

TECHNOMETRICS, ???? 0, VOL. 0, NO. 00
DOI 10.1198/004017008000000307

desirable. But the phase I analysis of multistage processes remains challenging and has not been thoroughly investigated in the literature.

One technical challenge is the use of multistage information in the analysis. Traditional methods developed for a single stage cannot fully use the information from a multistage operation where the input of the current process stage may be highly related to the output(s) of the early stage(s). Another challenge is the complexity of phase I analysis with uncertain multistage parameters. One task in phase I is to recognize the presence of OC conditions and identify which data (if any) are likely to be from the IC condition. But unlike in phase II monitoring, the parameters of the process in phase I analysis are not known, so that any strategy for phase I analysis depends on the simultaneous estimation of the parameters. Including OC data can lead to biased estimation, which in turn may affect the ability to detect the presence of the OC condition(s) (the "masking" problem). The situation may be more serious for multistage processes because of the intricacy of the model and the large number of parameters.

In this article we integrate the popular multivariate change point detection scheme with specific directional information based on a multistage state-space model. Based on this integration, we propose an effective multivariate change point method to assist in testing and estimating the sustained shift in a fixed multistage process sample by using multistage information. With a focus on the sustained shift case, we demonstrate that the proposed scheme uniformly outperforms existing approaches for multistage processes.

The remainder of the article is organized as follows. First, in Section 2 we describe the multistage process model in detail and review existing multivariate change point approaches that can be applied to multistage models. In Section 3 we consider a multivariate hypothesis test problem when a set of alternative hypotheses is available. Then we propose a directional multivariate change point method to resolve the unique challenges of multistage processes. We also give an appropriate approximation of the threshold for the test statistic. We analyze the test and estimation performance and compare it with the results of existing approaches in Section 4. Finally, in Section 5 we summarize the major contributions of the article and suggest issues for future research. We provide some necessary proofs and derivations in the Appendix.

2. MULTISTAGE MODELING AND TRADITIONAL CHANGE POINT APPROACHES

2.1 Multistage Modeling

Consider a manufacturing process composed of p stages. Without loss of generality, the stages are numbered in ascending order, such that if stage k precedes stage l, then $k < l$. A two-level linear state-space model generated from a practical application is usually used to describe the quality measurement at the kth stage of an IC process (Xiang and Tsung 2006),

$$\begin{aligned} y_k &= C_k x_k + v_k, \\ x_k &= A_k x_{k-1} + w_k, \end{aligned} \tag{1}$$

for $k = 1, \dots, p$, where x_k denotes the unobservable product quality characteristic. The first level of the model involves fitting the quality measurement to the quality characteristic, with C_k used to relate the unobservable process quality characteristic, x_k, to the observable quality measurement, y_k. The second level of the model involves modeling the transfer of the quality characteristic from the previous stage to the present stage, in which A_k denotes the transformation coefficient of the quality characteristic from stage $k-1$ to stage k. A_k and C_k for $k = 1, \dots, p$ are known constants (or matrixes) that are usually derived or estimated from engineering knowledge (see Jin and Shi 1999; Ding et al. 2002a for details). The process noise (say the common-cause variation and the unmodeled errors) is designated w_k, and the measurement error of the product quality is designated v_k.

Camelio, Hu, and Ceglarek (2003) used the foregoing state-space model to represent variation propagation of a multistage assembly system with compliant parts. Here x_k represents the part deviation at stage k, and y_k is the deviation vector containing all measurements at key product characteristic points. A_k is the dynamic matrix that represents the deviation change due to part transfer among stages and is determined jointly by relocation, deformation, and sensitivity matrixes. C_k is the observation matrix, corresponding to the information on the sensor number and sensor locations.

Such a linear state-space model structure provides an analytical engineering tool for modeling, analyzing, and diagnosing a multistage process. Extensive reviews of the state-space model have been given by Basseville and Nikiforov (1993) and Ding et al. (2002a). In this article we start from a univariate case, where A_k C_k, x_k, and y_k in model (1) are one-dimensional, and follow the model in the literature, $v_k \sim \mathrm{N}(0, \sigma_v^2)$, $w_k \sim \mathrm{N}(0, \sigma_{w_k}^2)$. The variance depends on the stage index, k, and the initial state, $x_0 \sim \mathrm{N}(a_0, \varepsilon^2)$. Usually in SPC applications, the unknown parameters, a_0, ε^2, σ_v^2, and $\sigma_{w_k}^2$, may be estimated through a large, good reference sample obtained in the phase I analysis.

Suppose that a multistage process data set contains m observations in the form $\{y_{1,j}, y_{2,j}, \dots, y_{p,j}\}$, where $y_{k,j}$ is a quality measurement observed from the jth product operated on the kth stage in the process described by model (1), $j = 1, 2, \dots, m$. The model based on the observations then can be written as

$$\begin{aligned} y_{k,j} &= C_k x_{k,j} + v_{k,j}, \\ x_{k,j} &= A_k x_{k-1,j} + w_{k,j}, \end{aligned} \tag{2}$$

where $v_{k,j}$ and $w_{k,j}$ are independent for $k = 1, \dots, p$ and $j = 1, 2, \dots, m$. Here we assume that $m > p + 1$. This assumption is not a restrictive one and can be easily satisfied in practical applications. It can ensure that the estimation of the covariance matrix in the change point approach is nonsingular, which we discuss later.

In this article we consider a typical change point problem in which a single step change occurs at some point, τ, in the mean of only one stage, say ζ; that is, the OC model can be represented as

$$\begin{aligned} y_{k,j} &= C_k x_{k,j} + v_{k,j}, \\ x_{k,j} &= A_k x_{k-1,j} + w_{k,j} + \delta \mathrm{I}_{\{k=\zeta, j>\tau\}}, \end{aligned} \tag{3}$$

where $k = 1, \dots, p$, $j = 1, 2, \dots, m$, and $\mathrm{I}_{\{\cdot\}}$ is the indicator function. We want to test the "no-change" null hypothesis, $H_0 : \delta = 0$, against the alternative, $H_A : \delta \neq 0$. We concentrate on this model assumption, which is common in industrial practice, in this article, and will investigate various extensions in future research.

2.2 Conventional Change Point Approaches

Conventionally, there are two possible approaches to develop testing methods for the multistage process change point problem. One approach is to conduct a test for the quality measurements of the product at the final stage (i.e., $y_{p,j}$) using classical univariate techniques, such as the parametric likelihood ratio test or the Schwarz information criterion (Csorgo and Horvath 1997; Chen and Gupta 2000). Obviously, using a single stage in a multistage process can be misleading and ineffective, because its measure is confounded with the cumulative effects from previous stages. Moreover, the impact of cross-correlation and autocorrelation due to the process dynamics and inertia of serially dependent stages should not be ignored. Xiang and Tsung (2006) have discussed the disadvantages of the online control scheme based on the observations of the final stage from the standpoint of efficient monitoring and diagnosing. Thus we believe that this method is not appropriate for testing and estimating the change point for the multistage model, and we will not proceed further with it.

Another conventional approach is to obtain the one-step forecast errors (OSFEs) (i.e., the residuals), based on the process model and then apply some traditional methods to them. For model (2), the standardized OSFEs have been given by Xiang and Tsung (2006) and are independently and identically distributed (iid) from the standard normal distribution (Durbin and Koopman 2001) under IC conditions and with known parameters. But in a phase I setting, the parameter values are unknown and must be estimated. Because of the complex form of the OSFEs and the possible presence of an OC condition, it appears to be very difficult (if not impossible) to obtain the analytical properties and make use of the OSFEs. Thus in this article we do not develop a test and estimation method for a multistage change point model based on OSFEs.

Alternatively, we may focus directly on original observations from each stage. As we discuss later, the method based on original observations can conveniently make full use of the underlying information in the multistage model. Thus we propose constructing a proper test for the multistage change point model by investigating the original product observations, $y_{1,j}, \dots, y_{p,j}$, directly.

Let $\mathbf{y}_j = (y_{1,j}, \dots, y_{p,j})'$ for $j = 1, 2, \dots, m$. Because the product observations are independent, we can regard $\mathbf{y}_1, \dots, \mathbf{y}_m$ as a sequence of independent p-dimensional multivariate normal vectors with unknown mean $\boldsymbol{\mu}_0$ and unknown variance matrix $\boldsymbol{\Sigma}$. Therefore, the test for a mean shift in the data set, $\{\mathbf{y}_j = (y_{1,j}, \dots, y_{p,j})', j = 1, 2, \dots, m\}$, may be considered a change point detection problem, which concerns the mean vector change under the assumption of an unknown but constant covariance matrix, when multivariate individual observations are available. Such a change point problem has been studied by Srivastava and Worsley (1986), who proposed a method (designated as the SW method hereinafter) to detect a change point in the mean vector. James, James, and Siegmund (1992) proposed a similar method to the SW method but with a different approximation of the critical value for the test statistics. Sullivan and Woodall (2000) proposed a preliminary analysis method for detecting a shift in the mean vector, the covariance matrix, or both, for multivariate individual observations and used the SW method as the benchmark. Related asymptotic results of the SW method also have been given by Chen and Gupta (2000). Because we are concerned mainly with the shift in the mean vector rather than the change in the variation, we use the SW method as a benchmark for comparisons with the proposed scheme in next section.

Here we briefly describe the SW method. Denote the means of the first l and last $m - l$ observations as $\bar{\mathbf{y}}_l = \frac{1}{l}\sum_{j=1}^{l} \mathbf{y}_j$, $\bar{\mathbf{y}}_{m-l} = \frac{1}{m-l}\sum_{j=l+1}^{m} \mathbf{y}_j$, and the pooled sample variance matrix as

$$\mathbf{W}_l = \frac{1}{m-2}\left\{\sum_{j=1}^{l}(\mathbf{y}_j - \bar{\mathbf{y}}_l)(\mathbf{y}_j - \bar{\mathbf{y}}_l)' + \sum_{j=l+1}^{m}(\mathbf{y}_j - \bar{\mathbf{y}}_{m-l})(\mathbf{y}_j - \bar{\mathbf{y}}_{m-l})'\right\}.$$

Obviously, under the OC model (3), $\mathbf{W}_l$ is an unbiased estimator of $\boldsymbol{\Sigma}$ only when $l \leq \tau < l + 1$. From an asymptotic viewpoint, it is consistent if $l/\tau = 1 + o(1)$; see the arguments in the proof of Proposition 2.

Then the standardized difference between the observations before and after the change point is

$$\mathbf{t}_l = \sqrt{\frac{l(m-l)}{m}}(\bar{\mathbf{y}}_l - \bar{\mathbf{y}}_{m-l}),$$

and the corresponding Hotelling T^2 statistic is

$$T_l^2 = \mathbf{t}_l'\mathbf{W}_l^{-1}\mathbf{t}_l, \qquad l = 1, \dots, m-1. \tag{4}$$

Then H_0 is rejected if $\max_{1 \leq l \leq m-1} T_l^2$ is large enough. The ML estimator for the location of change τ is given by

$$\widehat{\tau}_{SW} = \mathop{\arg\max}_{l=1,\dots,m-1} T_l^2. \tag{5}$$

The foregoing SW method is a general approach for testing and estimating the change point in multivariate observation cases. But using this method for a multistage process may cause us to overlook some important process information. We illustrate this point and propose our approach in the next section.

3. THE DIRECTIONAL MULTIVARIATE CHANGE POINT METHOD

Approaches to testing and estimating a change point, such as the SW method, are usually based on the two-sample likelihood ratio test by means of the binary segmentation procedure (Chen and Gupta 2000). Therefore, before studying the change point problem of the multistage model, we first consider corresponding one-sample multivariate hypothesis testing problems in Section 3.1. Then the analogous arguments can be readily extended to the two-sample multivariate testing problems with unknown parameters and lead to our proposed change point method for multistage processes.

3.1 Multivariate Hypothesis Testing Problems

Let $\mathbf{y}_j = (y_{1,j}, \dots, y_{p,j})'$, $j = 1, \dots, m$, be a random sample of size m from a p-variate normal population with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We consider the problem in testing the following hypothesis test:

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \leftrightarrow \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0.$$

For simplicity, we suppose that the covariance matrix, $\boldsymbol{\Sigma}$, is positive definite and known. Also, without loss of generality, we assume that $\boldsymbol{\mu}_0 = 0$. For this problem, the most popular test is the generalized likelihood ratio test (GLRT), for which the test statistic (Anderson 1984) is

$$m\bar{\mathbf{y}}_m' \boldsymbol{\Sigma}^{-1} \bar{\mathbf{y}}_m \sim \chi_p^2, \tag{6}$$

where χ_p^2 denotes the chi-squared distribution with p degrees of freedom. The GLRT has some optimal properties, including admissibility, minimax properties and being the optimal invariant test against H_1. When $\boldsymbol{\Sigma}$ is estimated, a Hotelling T^2 distribution should be used instead.

Sometimes we do have some knowledge about the alternative hypothesis, such as its specific direction, $\boldsymbol{\mu} = \delta \mathbf{d}$, where the direction, $\mathbf{d}$, is a known vector but the scale, δ, is an unknown constant. In such a case the hypothesis test would be

$$H_0 : \boldsymbol{\mu} = 0 \quad \leftrightarrow \quad H_1' : \boldsymbol{\mu} = \delta \mathbf{d}.$$

For this problem, the likelihood ratio test is considered to be the uniformly most powerful unbiased test (Lehmann 1991), which has a rejection region

$$m(\mathbf{d}'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 / \mathbf{d}'\boldsymbol{\Sigma}^{-1}\mathbf{d} > \chi_1^2(\alpha),$$

where α is the prespecified type I error and $\chi_1^2(\alpha)$ is the upper α percentile with a χ_1^2 distribution.

Furthermore, in a more practical situation, we may consider the following hypothesis test:

$$H_0 : \boldsymbol{\mu} = 0 \quad \leftrightarrow \quad H_1'' : \boldsymbol{\mu} = \delta \mathbf{d}_1 \text{ or } \boldsymbol{\mu} = \delta \mathbf{d}_2 \dots \text{ or } \boldsymbol{\mu} = \delta \mathbf{d}_r,$$

where the alternative hypothesis has several possible directions, $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_r$, with known vectors. To consider the case where r is finite, the GLRT will reject the null hypothesis if the test statistic,

$$\max_{i=1,\dots,r} \{m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 / \mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i\}, \tag{7}$$

is larger than a prespecified critical value.

To the best of our knowledge, there is no published theoretical optimal property of such a test method under the alternative hypothesis. However, the efficiency of this test can be expected to be superior to test (6) because of its more sharply focused rejection region. Moreover, when the null hypothesis is rejected, we naturally can use

$$\arg\max_{\mathbf{d}_i = \mathbf{d}_1, \dots, \mathbf{d}_r} \{m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 / \mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i\} \tag{8}$$

as an estimator of the true mean vector direction.

Next we present a proposition that establishes some good asymptotic properties of test (7) and estimator (8).

Proposition 1. For testing the null hypothesis, H_0, against the alternative, H_1'', the following hold:

a. For any finite set of directions, test (7) is asymptotically no less powerful than test (6).

b. Without loss of generality, we assume that $\boldsymbol{\mu} = \delta \mathbf{d}_k$, where $1 \le k \le r$. Then, the estimator of k,

$$\hat{k} = \arg\max_{1 \le i \le r} \{m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 / \mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i\},$$

is consistent.

We will find this proposition quite helpful in constructing our test and estimation of the change point in a multistage model in the following sections.

3.2 The Directional Multivariate Change Point Method

Here we propose a multivariate change point scheme that makes full use of the directional information from a multistage process. Suppose that a shift occurs at some stage, ζ, as represented by model (3). The expectation of $\mathbf{y}_j$ will change from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \delta \mathbf{d}_\zeta$, where $\mathbf{d}_\zeta = (d_{\zeta,1}, d_{\zeta,2}, \dots, d_{\zeta,p})'$, and

$$d_{\zeta,k} = \begin{cases} C_k \prod_{i=\zeta+1}^{k} A_i & \text{if } k \ge \zeta \\ 0 & \text{otherwise.} \end{cases}$$

Thus in the multistage process considered here, the process shift occurs only in one of p known directions $\mathbf{d}_\zeta$, $\zeta = 1, \dots, p$. This a priori shift direction information is a particular characteristic in multistage processes. If we could construct a multivariate change point scheme that makes use of this a priori knowledge properly and sufficiently, then the testing and estimation of a change point in the multistage process would be enhanced compared with the SW method based only on the GLRT, which does not take such constraining information into consideration.

The SW method combines a binary segmentation procedure with the two-sample mean test (Chen and Gupta 2000). On the other hand, the SW test statistic also can be obtained from the likelihood ratio procedure approach by assuming that $\boldsymbol{\Sigma}$ is known and then using an estimator to replace $\boldsymbol{\Sigma}$. Analogously, for the multistage model change point problem, by comparing testing statistics (6) and (7), it is natural to obtain the following test from (4): H_0 is rejected if

$$\max_{1 \le l \le m-1} U_l > c, \tag{9}$$

where

$$U_l = \max_{1 \le k \le p} (\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{t}_l)^2 / \mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k, \qquad l = 1, \dots, m-1, \tag{10}$$

and c is a constant for attaining the prespecified type I error probability α. This testing method integrates the binary segmentation procedure and the two-sample GLR mean test with directional information. Moreover, estimators of ζ and τ are given by

$$\widehat{\tau} = \arg\max_{1 \le l < m} \{U_l\} \tag{11}$$

and

$$\widehat{\zeta} = \arg\max_{1 \le k \le p} \{(\mathbf{d}_k'\mathbf{W}_{\widehat{\tau}}^{-1}\mathbf{t}_{\widehat{\tau}})^2 / \mathbf{d}_k'\mathbf{W}_{\widehat{\tau}}^{-1}\mathbf{d}_k\}. \tag{12}$$

Remark 1. Note that the foregoing directional multivariate change point (DMCP) test can be rewritten in the following form:

$$\max_{1\le k\le p}\left\{\max_{1\le l\le m-1}(\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{t}_l)^2/\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k\right\} > c.$$

On the other hand, the statistic $V_k \equiv \max_{1\le l\le m-1}(\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{t}_l)^2/\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k$ is similar to the test statistic based on the likelihood ratio approach for a change point problem set before and after the change point of the mean vector, $\boldsymbol{\mu}=\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}=\boldsymbol{\mu}_0+\delta\mathbf{d}_k$ (see the proof of Prop. 4 in the App.). The DMCP also can be seen as the combination of p special multivariate change point tests, each one being the simple hypothesis versus simple hypothesis. This can help us obtain an approximate significance level of the DMCP test, which we illustrate later.

Remark 2. Note that $\boldsymbol{\Sigma}$ can be represented as

$$\boldsymbol{\Sigma}=\begin{pmatrix}\operatorname{var}(y_{1,j}) & \operatorname{cov}(y_{1,j},y_{2,j}) & \cdots & \operatorname{cov}(y_{1,j},y_{p,j})\\ \operatorname{cov}(y_{1,j},y_{2,j}) & \operatorname{var}(y_{2,j}) & \cdots & \operatorname{cov}(y_{2,j},y_{p,j})\\ \vdots & \vdots & \vdots & \vdots\\ \operatorname{cov}(y_{1,j},y_{p,j}) & \operatorname{cov}(y_{2,j},y_{p,j}) & \cdots & \operatorname{var}(y_{p,j})\end{pmatrix},$$

where

$$\operatorname{var}(y_{k,j}) = C_k^2\left[\left(\prod_{i=1}^{k}A_i\right)^2\varepsilon^2+\sum_{l=1}^{k}\left(\prod_{i=l+1}^{k}A_i\right)^2\sigma_{w_l}^2\right]+\sigma_v^2$$

and

$$\operatorname{cov}(y_{l,j},y_{k,j}) = C_k\prod_{i=1}^{k}A_iC_l\prod_{i=1}^{l}A_i\varepsilon^2 + \sum_{m=1}^{l}C_k\prod_{i=m+1}^{k}A_iC_l\prod_{i=m+1}^{l}A_i\sigma_{w_m}^2, \qquad l<k.$$

Thus to estimate $\boldsymbol{\Sigma}$, we may estimate ε^2, σ_v^2, and $\sigma_{w_k}^2$ instead of using a pooled sample variance matrix. In fact, this problem will lead to the parameter estimation in a variance component model. The typical statistical estimation algorithms are ML estimation, restricted ML estimation (REML), and minimum norm quadratic unbiased estimation (MINQUE) (see Rao and Kleffe 1988). Zhou et al. (2004) suggested using MINQUE as an approximation of the ML estimate for multistage models because the computation load of MINQUE is much lower than that of ML estimation or REML. But we do not believe that MINQUE is suitable for phase I analysis, for two reasons. First, because MINQUE is obtained by solving $p+2$ nonlinear equations, nonconvergence of the iterative algorithm can occur. To reduce the frequency of nonconvergence, the use of good starting values is recommended (Rao 1997). But choosing such starting values is rather difficult in the presence of OC condition observations. In fact, the convergence of MINQUE may be questionable under the OC condition. Second, our change point method is based on the binary segmentation procedure, so that $m-1$ segments are needed. For each segment and each $\mathbf{d}_k$, the MINQUE of $\boldsymbol{\Sigma}$ must be obtained; that is, $p+2$ nonlinear equations are solved. When the sample size, m, and number of stages, p, are large, this is not a trivial computational task. Thus we suggest using the pooled sample variance matrix, $\mathbf{W}_l$, to estimate $\boldsymbol{\Sigma}$. This estimator is easy to calculate and has a closed form, so that we can obtain some analytical properties for the proposed DMCP.

Remark 3. It is noteworthy that for a multistage model (2), $\boldsymbol{\mu}_0$ can be represented as

$$\begin{aligned}\boldsymbol{\mu}_0 &= E(\mathbf{y}_j)\\ &= \left(C_1A_1a_0,\ldots,C_k\prod_{i=1}^{k}A_ia_0,\ldots,C_p\prod_{i=1}^{p}A_ia_0\right)\\ &\equiv a_0\boldsymbol{\varsigma};\end{aligned}$$

that is, in the p-variate vector, $\boldsymbol{\mu}_0$, there is only one unknown parameter, a_0. Note that we do not incorporate this point into the proposed test (9), because we assume that we have no prior knowledge about which part of the data set may be in control, that is, model (3). But sometimes we indeed have some additional information about which part of the data is in control. In such a case, without loss of generality, we assume that the observations before some unknown change point are IC, and it is equivalent to consider the test

$$H_0: E(\mathbf{y}_1)=E(\mathbf{y}_2)=\cdots=E(\mathbf{y}_m)=a_0\boldsymbol{\varsigma}$$

against the alternative

H_A: there is an integer τ such that

$$\begin{aligned}a_0\boldsymbol{\varsigma} &= E(\mathbf{y}_1)=\cdots=E(\mathbf{y}_\tau)\\ &\ne E(\mathbf{y}_{\tau+1})=\cdots=E(\mathbf{y}_m)\\ &= a_0\boldsymbol{\varsigma}+\delta\mathbf{d}_k\end{aligned}$$

for $k=1,\ldots,p$. The GLRT statistic will be

$$\begin{aligned}\max_{1\le k\le p}\Bigg\{&\max_{1\le l\le m-1}(m-l)\\ &\times(\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\bar{\mathbf{y}}_m\cdot\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\mathbf{d}_k-\mathbf{d}_k'\mathbf{W}_l^{-1}\bar{\mathbf{y}}_{m-l}\cdot\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\boldsymbol{\varsigma})^2\\ &\Bigg/\Bigg(\left[\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\boldsymbol{\varsigma}\cdot\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k-\frac{m-l}{m}(\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\mathbf{d}_k)^2\right]\\ &\times\boldsymbol{\varsigma}'\mathbf{W}_l^{-1}\boldsymbol{\varsigma}\Bigg)\Bigg\}.\end{aligned}$$

For testing the preceding change point hypothesis, applying this test statistic may improve the efficiency compared with the test (9).

To end this section, we present an asymptotic result on the consistency of the change point estimator (11) and the estimated stage where the shift occurs [eq. (12)].

Proposition 2. Suppose that $0<\lim_{m\to\infty}\tau/m=\theta<1$, where τ is the true change point in the observations. When the process is statistically OC as model (3) and the size of shift δ is a fixed constant, we then have the following:

a. $|\widehat{\tau}-\tau|=O_p(1)$, where $\widehat{\tau}$ is as defined in (11) and $O_p(1)$ means bounded in probability.

b. $\widehat{\zeta}\to\zeta$ in probability as $m\to\infty$, where $\widehat{\zeta}$ is as defined in (12).

This proposition ensures that the proposed estimators (11) and (12) are asymptotically effective. Although whether or not $\widehat{\tau}$ outperforms $\widehat{\tau}_{SW}$ remains undemonstrated analytically at this point in the article, the simulations in Section 4 show that $\widehat{\tau}$ indeed improves the accuracy of the change point estimates for the multistage processes considered in this article.

3.3 Approximate Critical Value

Srivastava and Worsley (1986) and James et al. (1992) presented two different methods for approximating the significance level of the SW test. In this section we study some properties of DMCP and then propose a suitable method to approximate its significance level or, equivalently, the critical value.

First, we give an approximation of $\Pr(V_k > c), k = 1, \dots, p$. The statistic T_l^2 in (4) is known to follow Hotelling's T^2 distribution, which does not depend on the parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$. To develop the approximations of $\Pr(V_k > c)$, we need to show that V_k has a similar affine-invariant property. Denoting the statistics as $(\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{t}_l)^2/\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k$ by $G_{l,k}$, we then have the following proposition.

Proposition 3. When a process is statistically IC, for any given vector $\mathbf{d}_k \neq 0$, the distribution of $G_{l,k}$ does not depend on $\mathbf{d}_k$, $\boldsymbol{\mu}_0$, or $\boldsymbol{\Sigma}$.

Based on the foregoing proposition, $\Pr(V_k > c)$ can be conveniently obtained by Monte Carlo simulations, because we can choose any set of values, $\mathbf{d}_k$, $\boldsymbol{\mu}_0$, or $\boldsymbol{\Sigma}$. Without loss of generality, let $\mathbf{d}_k' = (0, 0, \dots, 1) \equiv \mathbf{e}_p'$, $\boldsymbol{\mu}_0 = 0$, and $\boldsymbol{\Sigma}$ be the identity matrix. This proposition also makes approximating $\Pr(V_k > c), k = 1, \dots, p$ easy, because we need only study each specific choice for $\mathbf{d}_k$.

To approximate $\Pr(V_k > c)$, we need to explore the distributional properties of $G_{l,k}$. By Proposition 3, we have the following corollary.

Corollary 1. When the process is statistically IC, the statistic $G_{l,k}$ has the same distribution as

$$\frac{m-2}{m-p-1} F_1 \left(1 + \frac{p-1}{m-p} F_2\right),$$

where F_1 and F_2 are two independent random variables with F distributions with $(1,\ m-p-1)$ and $(p-1,\ m-p)$ degrees of freedom.

This corollary indicates that $G_{l,k}$ has asymptotically a chi-squared distribution with 1 degree of freedom as $m \to \infty$. In fact, when $\boldsymbol{\Sigma}$ is known, the random variable with $\boldsymbol{\Sigma}$ replacing $\mathbf{W}_l$ in $G_{l,k}$ will follow exactly the χ_1^2 distribution. Furthermore, by a modification of theorem 1.3.1 of Csorgo and Horvath (1997), we have the following proposition, which establishes the asymptotic null distribution of V_k.

Proposition 4. When the process is statistically IC, for every $k = 1, \dots, p$, we have

$$\lim_{m\to\infty} \Pr\{A(\log m)[V_k]^{1/2} \le t + D(\log m)\} = \exp(-2e^{-t}) \tag{13}$$

for all t, where $A(x) = (2\log x)^{1/2}$ and $D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi$.

We can use this proposition to get asymptotic critical values of V_k; however, in change point problems, the rate of convergence of the distribution of the test statistic based on binary segmentation is believed to be usually slow (see sec. 1.3 of Csorgo and Horvath 1997 for some discussions). Consequently, when m is not large enough, the approximation of (13) yields somewhat conservative results for small values of p. Moreover, empirically speaking, when the number of stages p is fairly large but m is small, using (13) may lead to a much larger type I error than the nominal value, due to inaccurate estimation of the high-dimensional nuisance parameter ($\boldsymbol{\Sigma}$). The result of some simulations that we conducted (not reported here) support these assertions.

In fact, the asymptotic distribution of V_k given by (13) equals that of the classical likelihood ratio, Z_m, for testing only a mean change (see Csorgo and Horvath 1997; Chen and Gupta 2000; and references therein). When there is no other unknown nuisance parameter, the critical value can be calculated using the recursion technique of Hawkins (1977). James et al. (1987, 1992) considered approximation through a d-dimensional Brownian bridge. A similar but more explicit suggestion was given by Csorgo and Horvath (1997, sec. 1.3); their recommended approximation is

$$\Pr(Z_m^{1/2} > x) \approx \frac{x\exp(-x^2/2)}{\sqrt{2\pi}} \left\{\ln(s) - \frac{1}{x^2}\ln(s) + \frac{4}{x^2}\right\}, \tag{14}$$

where $s = \frac{(1-h)(1-l)}{hl}$ and $h = l = (\ln(m))^{3/2}/m$. This is fairly accurate if x is not too small. Applying (14) to V_k will yield a large bias when m is not large enough, because the statistic $G_{l,k}$ is not exactly χ_1^2 distributed due to the variation in estimating $\boldsymbol{\Sigma}$; that is, it is usually necessary to take a larger critical value for V_k than what comes from (14). Later we describe a simulation study that which demonstrates this point. But the earlier interpretation of (13) suggests a heuristic approximation of $\Pr(V_k > c)$ and, consequently, an appropriate approximation of c, $\widehat{c}$. Given a desired type I error, α, say, we may first find a value c_1, defined as the value of x^2 where x solves (14) with the "$\approx$" replaced by "=." Then we can determine $\widehat{c}$ by the equation

$$F(\widehat{c}) \equiv \Pr(G_{l,k} < \widehat{c}) = F_{\chi_1^2}(c_1), \tag{15}$$

where $F_{\chi_1^2}(\cdot)$ is the cumulative distribution function (cdf) of the χ_1^2 distribution and $F(\cdot)$ is the cdf of $G_{l,k}$, which can be computed by numerical integration according to the distribution given in Corollary 1. (A Fortran program for calculating the value $\widehat{c}$ is available from the authors on request.) In contrast, $\Pr(V_k > c)$ can be approximated by first finding c_1 from (15). and then substituting $c_1^{1/2}$ for x in (14).

To illustrate the effectiveness of the proposed approximation method, in Table 1 we tabulate some simulated probabilities, $\Pr(V_k > \widehat{c})$, for various values of m and p. Here α is the desired type I error and α' is the estimated value from using $\widehat{c}$ estimated by 30,000 repetitions of a Monte Carlo simulation. It can be seen that the accuracy of the approximation, $\widehat{c}$, usually increases as m increases and decreases as p increases. The values of α' are close to the desired values, α, even with a small number of samples ($m = 20$) and a large number of stages ($p = 6$). Generally, $\widehat{c}$ yields a nice approximation in most cases.

Table 1. Performance of the approximation based on (14) and (15)

			$p=2$		$p=4$		$p=6$	
m	α	c_1	$\widehat{c}$	α'	$\widehat{c}$	α'	$\widehat{c}$	α'
20	.100	6.35	9.04	.121	12.2	.123	17.4	.130
	.050	7.89	11.9	.058	16.3	.059	23.5	.064
	.010	11.4	20.0	.010	28.3	.011	41.0	.011
50	.100	7.24	8.30	.110	9.11	.115	10.1	.110
	.050	8.82	10.3	.051	11.4	.053	12.6	.054
	.010	12.4	15.7	.010	17.3	.011	18.7	.012
100	.100	7.78	8.36	.106	8.72	.107	9.12	.109
	.050	9.36	10.2	.050	10.6	.051	11.1	.053
	.010	12.9	14.9	.011	15.5	.010	16.0	.010

Remark 4. As one of the referees pointed out, there may exist two or more solutions of (14) for some large values of m. Empirically speaking, we can simply take the largest solution as c_1 in such a situation. But if some very mild conditions are imposed on m and α, then we may obtain only one solution of (14). For example, suppose that $m \geq 10$ and $\alpha \leq .2$, which are very common in industrial practice and can be satisfied in most applications. By some direct calculations, we can show that the solution of (14) exists, is unique, and is within the range of $[2, \infty)$.

Finally, we turn to developing the approximation of the significance value of the DMCP test statistic $\max_{1\leq k\leq p} V_k$. The probability, $\Pr(\max_{1\leq k\leq p} V_k > c)$, seems rather difficult to derive exactly, because V_k, $k = 1, \ldots, p$, are usually positively correlated, and the correlations depend on the unknown parameter $\mathbf{\Sigma}$. A simple and natural approach to tackling this problem is to use the classical Bonferroni procedure, that is, $\Pr(\max_{1\leq k\leq p} V_k > c) \leq \sum_{k=1}^{p} \Pr(V_k > c)$. We would expect this to maintain a false alarm rate close to its nominal analog when the number of stages, p, is not large; nevertheless, this approximation is conservative for large p and lacks power if several V_k's are highly correlated.

Note that the form of the proposed DMCP test statistic is similar in spirit to that of a global test statistic in a multiple-testing problem. Thus here we recommend applying the well-known Simes modified Bonferroni procedure (see Simes 1986 for details) instead of the Bonferroni approximation. Given the strong empirical evidence of the modified procedure's superiority, researchers have begun using it in various multiple-testing applications. Formally, we summarize this procedure for the present context as follows:

Step 1. Compute the values of statistics V_k for $k = 1, \ldots, p$.
Step 2. Use (14) and (15) to obtain the corresponding approximate p values for each V_k, say $\widehat{P}_k$, $k = 1, \ldots, p$.
Step 3. Given a desired false alarm the rate α, reject null hypothesis if $\widehat{P}_{(i)} \leq i\alpha/p$ for at least one i, where $\widehat{P}_{(1)} \leq \cdots \leq \widehat{P}_{(p)}$ are the ordered values of $\widehat{P}_1, \ldots, \widehat{P}_p$.

Our extensive numerical simulations demonstrate that Simes's procedure works very well and usually provides less conservative approximations compared the Bonferroni method. Some simulation results are given in Section 4. But the theoretical conservativeness of this procedure for our considered test statistics remains unknown to us. In fact, the Simes method for testing the intersection of more than two hypotheses is known to control the probability of type I error when the underlying test statistics are independent (Simes 1986). Some efforts have been made to prove that the conservativeness of Simes's procedure may hold for dependent statistics with various multivariate distributions, such as those of Samuel-Cahn (1996), Sarkar and Chang (1997), and Sarkar (1998). All of these efforts rely on the assumption that the null distributions of test statistics satisfy some specific conditions. (See Sarkar 1998 for a thorough review and discussion.) For example, Sarkar (1998) considered a quite relaxed condition, say the multivariate totally positive of order two, which is satisfied by a large family of multivariate distributions. Nevertheless, the joint distribution of V_k, even the marginal distribution for each V_k (not the asymptotic one), is complicated and remains a challenge. Thus applying Sarkar's (1998) results to the DMCP test statistics seems very difficult. An ongoing effort of our group is to theoretically prove the conservativeness of Simes's procedure for the present multistage process problem.

4. PERFORMANCE COMPARISONS

In this section we first investigate the test and estimation performance of the proposed scheme through Monte Carlo simulation and compare the results with the SW method under various combinations of the multistage state-space parameters. We then conduct a sensitivity analysis to evaluate the effectiveness and robustness of the proposed scheme under misspecification of values of A_k and C_k.

4.1 Powers of the Two Tests

For simplicity, we consider only the case of type I error when $\alpha = .05$ and fix the number of stages to $p = 5$ and the number of observations to $m = 50$. Without loss of generality, we use the parameters $a_0 = 0$ and $\varepsilon^2 = \sigma_{w_k}^2 = \sigma_v^2 = 1$ throughout this section. We compare the proposed DMCP method with the alternative scheme, the SW method. As we know, the performance of the method for testing change points in a multistage process depends on the parameters A_k and C_k for $k = 1, 2, \ldots, p$; the different locations, ζ, where the shift initially occurs; the magnitude of the shift, δ; and the change point, τ. In this section, various levels of $\delta = .5, 1.0, 1.5, 2.0, 3.0, 4.0$ are considered. $(A_k, C_k) = (1.0, 1.0)$, $(A_k, C_k) = (1.2, .8)$, and $(A_k, C_k) = (.8, 1.2)$ for $\zeta = 1, 3, 5$ and $\tau = 20$ have been found to be consistent with the numerical comparison settings of Xiang and Tsung (2006).

Table 2 gives the powers of the tests obtained from 20,000 replications, with the critical values for the two methods listed in the last row. Note that these values are obtained by simulations, so the two methods can be fairly compared. The table also gives the simulated type I errors for the DMCP method, determined from the Simes and Bonferroni procedures.

Apparently, both the Simes and Bonferroni procedures perform well in controlling the false alarm rate, but the Simes approach indeed provides less conservative performance. With respect to the powers, Table 2 shows that the proposed DMCP is uniformly superior to the alternative SW method in detecting any magnitude of shifts. Also, it can be clearly seen that

Table 2. Performance comparisons between DMCP and SW with $\alpha = .05$, $\tau = 20$, and $m = 50$ for a five-stage process

	δ	$(A_k, C_k) = (1.0, 1.0)$ DMCP	$(A_k, C_k) = (1.0, 1.0)$ SW	$(A_k, C_k) = (1.2, .8)$ DMCP	$(A_k, C_k) = (1.2, .8)$ SW	$(A_k, C_k) = (.8, 1.2)$ DMCP	$(A_k, C_k) = (.8, 1.2)$ SW
$\zeta = 1$	.50	.075	.067	.068	.064	.081	.074
	1.00	.185	.157	.159	.134	.222	.185
	1.50	.435	.369	.362	.303	.528	.453
	2.00	.750	.670	.657	.569	.845	.782
	3.00	.993	.981	.979	.953	.999	.995
	4.00	1.000	1.000	1.000	1.000	1.000	1.000
$\zeta = 3$	.50	.081	.076	.069	.066	.084	.080
	1.00	.210	.183	.157	.135	.259	.223
	1.50	.509	.436	.370	.305	.622	.547
	2.00	.830	.757	.662	.577	.911	.864
	3.00	.998	.995	.980	.960	1.000	.999
	4.00	1.000	1.000	1.000	1.000	1.000	1.000
$\zeta = 5$	.50	.070	.067	.062	.059	.077	.075
	1.00	.175	.147	.119	.107	.226	.199
	1.50	.429	.370	.274	.235	.584	.505
	2.00	.758	.677	.522	.451	.890	.836
	3.00	.994	.983	.932	.888	1.000	.999
	4.00	1.000	1.000	.998	.996	1.000	1.000
Simes		.049		.048		.050	
Bonferroni		.046		.045		.045	
Critical value		17.1	24.3	17.0	24.3	17.0	24.3

the DMCP method's superiority over its counterpart still holds when the values of the parameters, A_k, C_k, and ζ, vary. Here we do not tabulate the performance comparisons of these two methods for shifts at other change points, because similar results were obtained.

It is worth noting that the difference in performance between the DMCP and SW methods will be more prominent in higher-dimensional problems. To verify this point, we consider a multistage process of $p = 20$ and fix $\zeta = 10$, $\tau = 40$, and $m = 100$ for simplicity. The other parameters are the same as those given in Table 2. Table 3 gives the simulation results. Compared with the power results given in Table 2, the superiority of the DMCP method is more remarkable in the case where $p = 20$, especially for moderate magnitudes of change, say $\delta = 1.5$ or 2.0.

Based on the results given in Tables 2 and 3 and other simulations of various multistage model parameters (available from the authors on request), we conclude that the DMCP method, by taking advantage of incorporating shift direction information, is almost always better than the SW method in detecting mean process change. In the following section, we assess the effectiveness of the corresponding estimation methods as well.

4.2 Performance of the Estimators (11) and (12)

Here we investigate the performance of the proposed approach in estimating the change point after the null hypothesis is rejected. We compare the performance of two change point estimators, $\widehat{\tau}_{SW}$ in (5) and $\widehat{\tau}$ in (11). The number of stages, $p = 5$, and the number of observations, $m = 50$, are considered again. A total of 50,000 independent series are generated in the simulations. Note that any series for which no signal was trigged is discarded. The process change point is simulated at $\tau = 10$ and $\tau = 20$. Again the parameter combinations

Table 3. Performance comparisons between DMCP and SW with $\alpha = .05$, $\tau = 40$, and $m = 100$, and $\zeta = 10$ for a 20-stage process

δ	$(A_k, C_k) = (1.0, 1.0)$ DMCP	$(A_k, C_k) = (1.0, 1.0)$ SW	$(A_k, C_k) = (1.2, .8)$ DMCP	$(A_k, C_k) = (1.2, .8)$ SW	$(A_k, C_k) = (.8, 1.2)$ DMCP	$(A_k, C_k) = (.8, 1.2)$ SW
.50	.071	.071	.067	.067	.070	.078
1.00	.254	.174	.173	.128	.254	.225
1.50	.710	.480	.520	.325	.710	.613
2.00	.968	.856	.872	.661	.968	.942
3.00	1.000	.999	1.000	.992	1.000	1.000
4.00	1.000	1.000	1.000	1.000	1.000	1.000
Simes	.048		.047		.048	
Bonferroni	.043		.041		.042	
Critical value	27.5	63.6	27.4	63.6	27.4	63.6

TECH asa v.2007/01/31 Prn:25/07/2008; 15:23 F:tech07057.tex; (Roberta) p. 9

Table 4. The averages (AVE), standard deviations (SD), and precisions of change point estimates for shift at the third stage in a five-stage process

	AVE		SD		P_1		P_2		P_3		
δ	$\widehat{\tau}$	$\widehat{\tau}_{SW}$	$\widehat{\tau}$	$\widehat{\tau}_{SW}$	$\widehat{\tau}$	$\widehat{\tau}_{SW}$	$\widehat{\tau}$	$\widehat{\tau}_{SW}$	$\widehat{\tau}$	$\widehat{\tau}_{SW}$	P_ζ
$\tau = 10$											
1.00	16.7	17.1	12.8	13.1	.27	.23	.37	.33	.45	.41	.63
2.00	11.3	11.7	5.9	6.6	.61	.56	.73	.69	.81	.76	.93
3.00	10.3	10.4	2.5	3.0	.82	.79	.90	.88	.94	.93	.99
4.00	10.1	10.1	1.1	1.2	.93	.92	.97	.96	.99	.98	1.00
$\tau = 20$											
1.00	21.6	21.8	10.0	10.6	.27	.24	.37	.33	.44	.41	.75
2.00	20.3	20.3	4.2	4.9	.61	.57	.73	.70	.81	.77	.97
3.00	20.0	20.0	1.7	2.0	.82	.81	.91	.89	.95	.94	1.00
4.00	20.0	20.0	.9	.9	.93	.93	.97	.97	.99	.99	1.00

$(A_k, C_k) = (1.0, 1.0)$ and $\zeta = 3$ are considered. To quantify the diagnostic precision of the estimators, in addition to calculating the averages and the standard deviations of the change point estimates, Table 4 gives the probabilities, $\Pr(|\widehat{\tau} - \tau| \le 1)$, $\Pr(|\widehat{\tau} - \tau| \le 2)$, and $\Pr(|\widehat{\tau} - \tau| \le 3)$ (denoted as P_1, P_2, and P_3).

For a small shift of 1.0, the two estimators both appear to be biased in estimating the process change point, especially in the case where $\tau = 10$. The proposed $\widehat{\tau}$ performs uniformly better in terms of the average and standard deviation for any magnitude of shift. Moreover, $\widehat{\tau}$ has better performance in terms of diagnostic precision. Other extensive simulations conducted for various parameter combinations also demonstrated that $\widehat{\tau}$ appears to be more accurate in estimating τ. In addition, the probabilities $\Pr(\widehat{\zeta} = \zeta)$ (P_ζ), given in the last column of Table 4, also indicate that the estimates of ζ (12) are fairly accurate.

4.3 Sensitivity Analysis

In general, more powerful model-based strategies come at the price of being more adversely affected by deviations from the model assumptions. The performance of a model-based method depends on the nature and degree of the misspecification of the assumption. As mentioned in Section 1, A_k and C_k usually are derived or estimated from engineering knowledge. In practical applications, we may have inaccurate estimations of these parameters, especially if there are many stages. Estimation errors in A_k and C_k will yield incorrect direction vectors and adversely affect the performance of the DMCP method. Here we conduct some simulations to check how the DMCP method performs when A_k and C_k are misspecified.

For simplicity, here we consider $\alpha = .05$, $p = 20$, $\zeta = 10$, $\tau = 40$, and $m = 100$, the same values as in Table 3. Suppose that the true values of A_k and C_k are fixed as 1.0. Let $\widehat{A}_{k^*}$ and $\widehat{C}_{k^*}$ denote the incorrect values for the stage k^*. We consider two situations, (I) $k^* = 5$, 10, and 15 and (II) $k^* = 2i$ for $i = 1, 2, \ldots, 10$, which correspond to low and high degrees of parameter misspecification. For each situation, the following $\widehat{A}_{k^*}$ and $\widehat{C}_{k^*}$ are considered: (a) $\widehat{A}_{k^*} = .8$, (b) $\widehat{C}_{k^*} = .8$, (c) $\widehat{A}_{k^*} = 1.2$, (d) $\widehat{C}_{k^*} = 1.2$, (e) $\widehat{A}_{k^*} = 1.2$ and $\widehat{C}_{k^*} = .8$, (f) $\widehat{A}_{k^*} = .8$ and $\widehat{C}_{k^*} = 1.2$, (g) $\widehat{A}_{k^*} = .8$ and $\widehat{C}_{k^*} = .8$, and (h) $\widehat{A}_{k^*} = 1.2$ and $\widehat{C}_{k^*} = 1.2$. Table 5 gives the powers of the tests, obtained from 20,000 replications. Note that the values in the third and fourth columns of the table are from Table 3, where the values of A_k and C_k are prespecified correctly.

We can see that when the degree of misspecification is low, such as in (I), the DMCP approach still generally performs better than the SW method except for very small δ, but its superiority is reduced compared with the power performance in the fourth column. For the situation (II), where many parameters are estimated incorrectly, as we would expect, misspecification of A_k and C_k has a significantly adverse effect on the performance of the DMCP approach, especially in the last case, (h).

Table 5. Performance comparison with misspecified parameters A_k and C_k

			DMCP	DMCP with misspecified values of A_k and C_k							
	δ	SW	true	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
(I)	.50	.071	.071	.070	.068	.071	.068	.070	.069	.071	.069
	1.00	.174	.254	.242	.218	.248	.238	.220	.242	.215	.221
	1.50	.480	.710	.686	.642	.690	.672	.639	.679	.600	.615
	2.00	.856	.968	.961	.938	.958	.954	.933	.957	.919	.927
	3.00	.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
(II)	.50	.071	.071	.068	.066	.066	.067	.065	.067	.067	.061
	1.00	.174	.254	.220	.191	.168	.215	.160	.224	.189	.124
	1.50	.480	.710	.627	.549	.481	.618	.471	.639	.510	.309
	2.00	.856	.968	.937	.892	.823	.929	.816	.937	.860	.608
	3.00	.999	1.000	1.000	1.000	.998	1.000	.994	1.000	1.000	.959

Thus although the proposed approach can still work quite well even when values of A_k and C_k are misspecified, we believe that it is critical to estimate A_k and C_k accurately when using a model-based method.

5. CONCLUSIONS

Phase I analysis of multistage processes remains a challenging problem that has not been thoroughly investigated in the literature. In this article we have focused on the problem of testing and estimating the change point in a phase I reference sample. We have proposed an approach that combines the classical binary segmentation procedure and the multivariate two-sample test. This novel approach fully incorporates the directional information based on the multistage state-space model for testing the stability of multistage processes. We give an appropriate approximation of the threshold for the test statistic. We also explore an estimation approach to identify an OC stage and locate the change point, with the estimators shown to be consistent. Our simulation results also show that the proposed scheme consistently outperforms conventional multivariate approaches to multistage processes, unless the process model is grossly misspecified.

Note that although the focus of this article is on a univariate process, many multistage processes are actually multivariate in practice. Extension of our proposed DMCP scheme to a multivariate multistage process warrants further research. Moreover, it should be pointed out that our proposed approach can be readily extended to detect multiple change points by using the binary segmentation method recursively (Vostrikova 1981; Yao 1988).

ACKNOWLEDGMENTS

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article. This work was completed when Zou was a research assistant at the Hong Kong University of Science and Technology, whose hospitality is appreciated and acknowledged. This research was supported by RGC Competitive Earmarked Research grants 620606, National Science Foundation of Tianjin grant 07JCYBJC04300, and **???** (NNSF) of China Grant 10771107. Liu's work also was supported by NNSF of China grant 10571093 and SRFDP of China grant 20050055038.

APPENDIX: PROOFS

Proof of Proposition 1

a. Under the null hypothesis, we have

$$\Pr\{m\bar{\mathbf{y}}_m'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m > C'\} = \alpha$$

and

$$\Pr\Big\{\max_{i=1,\dots,r}\{m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2/\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i\} > C\Big\} = \alpha,$$

where C' is the upper α percentile of the χ_p^2 distribution and C is some chosen constant. Here we denote $m\bar{\mathbf{y}}_m'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m$ as $V(p)$ and $m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2/\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i$ as V_i, which follow a χ_1^2 distribution under H_0. As is well known, the GLRT is equivalent to

$$\max_{i=1,\dots,\infty}\{m(\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2/\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i\} > C',$$

where $\{d_i, i = 1,\dots,\infty\}$ is a countable set dense in the parameter space (Roy 1953). It immediately follows that $C' > C$.

Without loss of generality, we suppose that $\boldsymbol{\mu} = \delta\mathbf{d}_1$ and $\delta > 0$ under H_1''. It is obvious that these two multivariate tests are consistent and unbiased for testing H_1''. Note that under H_1'', $\Pr\{\max_{i=1,\dots,r} V_i > C\} \ge \Pr\{V_1 > C\}$. Thus it suffices to prove that

$$\Pr\{V_1 < C\} \le \Pr\{V(p) < C'\} \to 0 \quad \text{as } m \to \infty,$$

where $V_1 \sim \chi_1^2(m\delta^2)$ and $V(p) \sim \chi_p^2(m\delta^2)$. Here $\chi_p^2(m\delta^2)$ is the noncentral chi-squared distribution with the noncentral parameter $(m\delta^2)^{1/2}$. For simplicity, denote $\xi = \sqrt{m}\delta \to \infty$ as $m \to \infty$. Note that $V(p)$ can be partitioned into $U(p-1)$ and $V(1)$, which are distributed as χ_{p-1}^2 and $\chi_1^2(\xi^2)$. Consequently, by $C' > C$, there exists $\varepsilon > 0$, so that $C' - (C+\varepsilon) > 0$. We then have

$$\begin{aligned}
&\frac{\Pr\{V(p) < C'\}}{\Pr\{V_1 < C\}}\\
&\quad = \frac{\Pr\{U(p-1) + V(1) < C'\}}{\Pr\{V_1 < C\}}\\
&\quad \ge \frac{\Pr\{V(1) < C+\varepsilon\}}{\Pr\{V_1 < C\}}\Pr\{U(p-1) < C' - (C+\varepsilon)\}\\
&\quad \equiv \triangle.
\end{aligned}$$

The inequality derives from the fact that $U(p-1)$ and $V(1)$ are independent. Then, using the fact that $1 - \Phi(x) \approx \frac{\phi(x)}{x}$ [where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cdf and density function of standard normal distribution] for large x, we have

$$\begin{aligned}
&\frac{\Pr\{V(1) < C+\varepsilon\}}{\Pr\{V_1 < C\}}\\
&\quad = \frac{\Pr\{[z+\xi]^2 < C+\varepsilon\}}{\Pr\{[z+\xi]^2 < C\}}\\
&\quad \approx \frac{1 - \Phi(\xi - \sqrt{C+\varepsilon})}{1 - \Phi(\xi - \sqrt{C})}\\
&\quad \approx \exp\left\{(\sqrt{C+\varepsilon} - \sqrt{C})\xi - \frac{\varepsilon}{2}\right\} \quad \text{as } m \to \infty,
\end{aligned}$$

where z is a standard normal random variable.

By noting that the probability $\Pr\{U(p-1) < C' - (C+\varepsilon)\}$ is a value greater than 0 for the fixed ε, it follows that

$$\begin{aligned}
\triangle &\approx \exp\left\{(\sqrt{C+\varepsilon} - \sqrt{C})\xi - \frac{\varepsilon}{2}\right\}\\
&\quad \times \Pr\{U(p-1) < C' - (C+\varepsilon)\}\\
&\to \infty \quad \text{as } m \to \infty.
\end{aligned}$$

b. Without loss of generality, we assume that $\delta > 0$. For notational convenience, let $\mathbf{u}_i = \mathbf{d}_i/\sqrt{\mathbf{d}_i'\boldsymbol{\Sigma}^{-1}\mathbf{d}_i}$, $i = 1,\dots,r$. Then the estimator (8) can be rewritten as

$$\arg\max_{\mathbf{u}_i=\mathbf{u}_1,\dots,\mathbf{u}_r}\{m(\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2\}.$$

Thus proving the consistency of this estimator is equivalent to showing that

$$\Pr\left\{\bigcup_{i\neq k}[(\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 < (\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2]\right\} \to 0 \quad \text{as } m\to\infty.$$

Because

$$\Pr\left\{\bigcup_{i\neq k}[(\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 < (\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2]\right\} \le \sum_{i\neq k}\Pr\{(\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 < (\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2\},$$

it suffices to prove that

$$\sum_{i\neq k}\Pr\{(\mathbf{u}_j'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 - (\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 > 0\} \to 0 \quad \text{as } m\to\infty.$$

Denote $z_i = (\mathbf{u}_j'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2 - (\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)^2$. For any $\varepsilon > 0$,

$$\begin{aligned}\sum_{i\neq k}&\Pr\{z_i - E(z_i) + E(z_i) > \varepsilon\}\\ &= \sum_{i\neq k}\Pr\left\{\left[z_i - E(z_i) - \frac{\varepsilon}{2}\right] + \left[E(z_i) - \frac{\varepsilon}{2}\right] > 0\right\}\\ &\le \sum_{i\neq k}\Pr\left\{z_i - E(z_i) > \frac{\varepsilon}{2}\right\} + \sum_{i\neq k}\Pr\left\{E(z_i) > \frac{\varepsilon}{2}\right\}. \qquad (\text{A.1})\end{aligned}$$

Under $\boldsymbol{\mu} = \delta\mathbf{d}_k$, we have that $\bar{\mathbf{y}}_m \sim \mathrm{N}_p(\delta^*\mathbf{u}_k, \frac{1}{m}\boldsymbol{\Sigma})$, where $\delta^* = \delta(\mathbf{d}_k'\boldsymbol{\Sigma}^{-1}\mathbf{d}_k)^{1/2}$. It is not difficult to verify that $z_i \to E(z_i)$ as $m\to\infty$ and $\mathrm{var}(z_i) = O(m^{-1})$. Thus, by the Chebychev inequality, the first term of (A.1) tends 0 as $m\to\infty$.

On the other hand, $\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\mathbf{u}_i = 1$, $i = 1,\ldots,r$, yield

$$\begin{aligned}E(z_i) &= [E(\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)]^2 - [E(\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\bar{\mathbf{y}}_m)]^2\\ &= \delta^*[(\mathbf{u}_i'\boldsymbol{\Sigma}^{-1}\mathbf{u}_k)^2 - (\mathbf{u}_k'\boldsymbol{\Sigma}^{-1}\mathbf{u}_k)^2].\end{aligned}$$

From $(\mathbf{u}_k - \mathbf{u}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{u}_k - \mathbf{u}_i) \ge 0$ and $(\mathbf{u}_k + \mathbf{u}_i)'\boldsymbol{\Sigma}^{-1}(\mathbf{u}_k + \mathbf{u}_i) \ge 0$, we have that $E(z_i) < 0$. It immediately follows that the second term of (A.1) equals 0. Therefore, (A.1) tends to 0 as $m\to\infty$, which completes the proof.

Proof of Proposition 2

a. Bhattacharya (1987) studied the asymptotic distribution of the ML estimate of the change point in a general multiparameter case. Gombay and Horvath (1996) also considered the asymptotic behavior of the ML estimate of a change point in a exponential family case. For the sake of simplicity, they assumed no nuisance parameter (Csorgo and Horvath 1997, sec. 1.5). Because our suggested testing statistics are somewhat "special," the methods in the preceding articles cannot be directly extended to prove the consistency of the estimator $\widehat{\tau}$. Here we give a short proof.

We first consider $\tau < l < m$. Assume that when $m\to\infty$, $l/m \to \theta_1$, where $\theta_1 \in (0,1)$ and $\theta_1 > \theta$. By the law of large numbers, we can obtain

$$\begin{aligned}m^{-1/2}\mathbf{t}_l &= \delta\theta\sqrt{\theta_1^{-1} - 1}\,\mathbf{d}_\zeta + O_p(m^{-1/2}),\\ m^{-1/2}\mathbf{t}_\tau &= \delta\sqrt{\theta(1-\theta)}\,\mathbf{d}_\zeta + O_p(m^{-1/2}),\\ \mathbf{W}_\tau &= \boldsymbol{\Sigma} + O_p(m^{-1/2}),\end{aligned}$$

and

$$\mathbf{W}_l = \boldsymbol{\Sigma}_1 + O_p(m^{-1/2}),$$

where $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma} + \theta(1 - \theta\theta_1^{-1})\mathbf{d}_\zeta\mathbf{d}_\zeta'$. Consequently, it is easy to verify that $m^{-1}G_{\tau,\zeta} = O_p(1) > 0$ and $m^{-1}G_{l,k} = O_p(1)$. Thus to prove part a, it is sufficient to show that when $m\to\infty$, for any $1\le k\le p$, $\tau < l < m$, $m^{-1}(G_{\tau,\zeta} - G_{l,k}) > 0$, a.e. It then follows from the definition that

$$\begin{aligned}m^{-1}&(G_{\tau,\zeta} - G_{l,k})\\ &= \delta^2\left[\frac{\theta(1-\theta)(\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)^2}{\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta} - \frac{(\theta_1^{-1}-1)\theta^2(\mathbf{d}_\zeta'\boldsymbol{\Sigma}_1^{-1}\mathbf{d}_k)^2}{\mathbf{d}_k'\boldsymbol{\Sigma}_1^{-1}\mathbf{d}_k}\right]\\ &\quad\times\bigl(1 + O_p(m^{-1/2})\bigr)\\ &= \delta^2\frac{(\theta_1^{-1}-1)\theta^2}{\mathbf{d}_k'\boldsymbol{\Sigma}_1^{-1}\mathbf{d}_k}\mathbf{d}_k'\boldsymbol{\Sigma}_2\mathbf{d}_k\bigl(1 + O_p(m^{-1/2})\bigr),\end{aligned}$$

where $\boldsymbol{\Sigma}_2 = \gamma\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_1^{-1}\mathbf{d}_\zeta\mathbf{d}_\zeta'\boldsymbol{\Sigma}_1^{-1}$ and $\gamma = (\theta^{-1} - 1)/(\theta_1^{-1} - 1) > 1$. Now we just need to show that the matrix $\boldsymbol{\Sigma}_2$ is positive definite, which can be seen from

$$\begin{aligned}\boldsymbol{\Sigma}_2 &= (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)\boldsymbol{\Sigma}_1^{-1}[\gamma\boldsymbol{\Sigma}_1 - (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)^{-1}\mathbf{d}_\zeta\mathbf{d}_\zeta']\boldsymbol{\Sigma}_1^{-1}\\ &> (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)\boldsymbol{\Sigma}_1^{-1}[\boldsymbol{\Sigma} - (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)^{-1}\mathbf{d}_\zeta\mathbf{d}_\zeta']\boldsymbol{\Sigma}_1^{-1}\\ &= (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}^{1/2}\\ &\quad\times\bigl[\mathbf{I} - (\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{d}_\zeta\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1/2}\bigr]\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}_1^{-1}\\ &\ge 0,\end{aligned}$$

where the first inequality comes from $\boldsymbol{\Sigma}_1 - \boldsymbol{\Sigma} = \theta(1 - \theta\theta_1^{-1})\mathbf{d}_\zeta\mathbf{d}_\zeta' \ge 0$ and the second inequality comes from the matrix $(\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1}\mathbf{d}_\zeta)^{-1}\boldsymbol{\Sigma}^{-1/2}\mathbf{d}_\zeta\mathbf{d}_\zeta'\boldsymbol{\Sigma}^{-1/2}$ being symmetric and idempotent. Here for two symmetric matrixes, $\mathbf{C}$ and $\mathbf{D}$, $\mathbf{C}\ge\mathbf{D}$ means that $\mathbf{C}-\mathbf{D}$ is semipositive definite. The proof for $1 < l < \tau$ is analogous to the foregoing arguments and is omitted here.

b. By $|\widehat{\tau} - \tau| = O_p(1)$ and a similar argument as in the proof of Proposition 1.b, the proof of consistency is straightforward and is omitted here.

Proof of Proposition 3

First, let $\mathbf{Z}_i = \boldsymbol{\Sigma}^{-1/2}\mathbf{y}_i$, $i = 1,\ldots,m$. Then it follows that

$$G_{l,k} = (\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{t}_l)^2/\mathbf{d}_k'\mathbf{W}_l^{-1}\mathbf{d}_k = (\tilde{\mathbf{d}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{t}}_l)^2/\tilde{\mathbf{d}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{d}}_k,$$

where $\tilde{\mathbf{d}}_k = \boldsymbol{\Sigma}^{-1/2}\mathbf{d}_k$ and $\tilde{\mathbf{W}}_l$ and $\tilde{\mathbf{t}}_l$ are given by substituting $\mathbf{Z}_i$ into $\mathbf{W}_l$ and $\mathbf{t}_l$ instead of $\mathbf{y}_i$. Obviously, $\mathbf{Z}_i$, $i = 1,\ldots,m$, is a random sample of size m from a p-variate standard normal population. Thus $\tilde{\mathbf{t}}_l$ and $(m-2)\tilde{\mathbf{W}}_l$ are independently distributed as a p-variate standard multivariate normal distribution and a Wishart distribution with $m-2$ degrees of freedom. Thus it suffices to show that the distribution of $G_{l,k}$ does not depend on $\tilde{\mathbf{d}}_k$, which is some unknown vector.

It can be easily seen that $G_{l,k}$ is scale-invariant for $\tilde{\mathbf{d}}_k$. We show that $G_{l,k}$ has an orthogonal transformation-invariant property. Let $\mathbf{e}_k = \boldsymbol{\Gamma}\tilde{\mathbf{d}}_k$ and $\mathbf{U}_l = \boldsymbol{\Gamma}\tilde{\mathbf{W}}_l\boldsymbol{\Gamma}'$, where $\boldsymbol{\Gamma}$ is a orthogonal matrix. We need to show that $(\tilde{\mathbf{d}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{t}}_l)^2/\tilde{\mathbf{d}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{d}}_k$ and $(\tilde{\mathbf{e}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{t}}_l)^2/\tilde{\mathbf{e}}_k'\tilde{\mathbf{W}}_l^{-1}\tilde{\mathbf{e}}_k$ are identically distributed. Because

$$\frac{(\mathbf{e}_k'\tilde{\mathbf{W}}_l^{-1}\mathbf{t}_l)^2}{\mathbf{e}_k'\tilde{\mathbf{W}}_l^{-1}\mathbf{e}_k} = \frac{(\tilde{\mathbf{d}}_k'\mathbf{U}_l^{-1}\boldsymbol{\Gamma}\tilde{\mathbf{t}}_l)^2}{\tilde{\mathbf{d}}_k'\mathbf{U}_l^{-1}\tilde{\mathbf{d}}_k}$$

and $\boldsymbol{\Gamma}\tilde{\mathbf{t}}_l$ is independent of $\mathbf{U}_l$ and follows a standard multivariate normal distribution, we can complete the proof by showing that $\mathbf{U}_l$ and $\tilde{\mathbf{W}}_l$ are identically distributed.

As we know, the density function of $\tilde{\mathbf{W}}_l$ is $f_{\tilde{\mathbf{W}}_l} \propto |\tilde{\mathbf{W}}_l|^{-(m-p-2)/2}\exp\{-\frac{1}{2}\mathrm{tr}(\tilde{\mathbf{W}}_l)\}$, and the Jacobian transformation is $J(\mathbf{U}_l \to \tilde{\mathbf{W}}_l) = |\boldsymbol{\Gamma}|^{p+1} = 1$, Then the density function of $\mathbf{U}_l$ is

$$\begin{aligned} f_{\mathbf{U}_l} &\propto |\tilde{\mathbf{W}}_l|^{-(m-p-2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\tilde{\mathbf{W}}_l)\right\}J(\mathbf{U}_l \to \tilde{\mathbf{W}}_l)^{-1} \\ &= |\boldsymbol{\Gamma}\mathbf{U}_l\boldsymbol{\Gamma}'|^{-(m-p-2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Gamma}\mathbf{U}_l\boldsymbol{\Gamma}')\right\} \\ &= |\mathbf{U}_l|^{-(m-p-2)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{U}_l)\right\}, \end{aligned}$$

which means that $\mathbf{U}_l$ and $\tilde{\mathbf{W}}_l$ are identically distributed.

Proof of Corollary 1

By Proposition 3, without loss of generality, we take $\mathbf{d}_k' = (0, 0, \ldots, 1) \equiv \mathbf{e}_p'$, $\boldsymbol{\mu}_0 = 0$, and $\boldsymbol{\Sigma}$ to be the identity matrix. Also, for simplicity, here we suppress the symbol "l" in $\mathbf{W}_l^{-1}$ and $\mathbf{t}_l$. Then

$$\begin{aligned} G_{l,k} &= (\mathbf{e}_p'\mathbf{W}^{-1}\mathbf{t})^2/\mathbf{e}_p'\mathbf{W}^{-1}\mathbf{e}_p \\ &= \mathbf{t}'[\mathbf{W}^{-1}\mathbf{e}_p\mathbf{e}_p'\mathbf{W}^{-1}/\omega^{pp}]\mathbf{t} \equiv \mathbf{t}'\mathbf{V}\mathbf{t}, \end{aligned}$$

where ω^{ij} is the (i,j) element of $\mathbf{W}^{-1}$. Obviously, $\mathbf{V}$ is a semi-positive definite matrix with rank 1. Thus there exists an orthogonal matrix, $\boldsymbol{\Gamma}$, such that $\boldsymbol{\Gamma}\mathbf{V}\boldsymbol{\Gamma}' = \mathrm{diag}(\mathbf{0}, \mathbf{0}, \ldots, \mathrm{tr}(\mathbf{V}))$. Then $\boldsymbol{\Gamma}\mathbf{t}|\boldsymbol{\Gamma} \sim \mathrm{N}_p(0, \mathbf{I})$ and is independent of $\boldsymbol{\Gamma}$. Consequently, the unconditioned distribution of $\boldsymbol{\Gamma}\mathbf{t} \equiv \tilde{\mathbf{t}}$ is also a standard multivariate normal distribution. Therefore, $G_{l,k}$ and $\tilde{t}_p^2 \cdot \mathrm{tr}(\mathbf{V})$ are identically distributed. In addition, we can see that $\tilde{\mathbf{t}}$ is independent of $\mathrm{diag}(0, 0, \ldots, \mathrm{tr}(\mathbf{V}))$ by noting that $\mathbf{t}$ is independent of $\mathbf{W}^{-1}$. We partition $\mathbf{W}$ as $\begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & W_{22} \end{pmatrix}$ and $\mathbf{W}^{-1}$ as $\begin{pmatrix} \mathbf{W}^{11} & \mathbf{W}^{12} \\ \mathbf{W}^{21} & \omega^{pp} \end{pmatrix}$, where $\mathbf{W}_{11}$ and $\mathbf{W}^{11}$ are $(p-1)$-dimensional matrixes. Consequently, we have

$$\begin{aligned} \tilde{t}_p^2 \cdot \mathrm{tr}(\mathbf{V}) &= \tilde{t}_p^2\left(\frac{\mathbf{W}^{21}\mathbf{W}^{12}}{\omega^{pp}} + \omega^{pp}\right) \\ &= \tilde{t}_p^2\left(\frac{(\omega^{pp})^2\mathbf{W}_{21}\mathbf{W}_{11}^{-2}\mathbf{W}_{12}}{\omega^{pp}} + \omega^{pp}\right) \\ &= \tilde{t}_p^2\omega^{pp}(\mathbf{W}_{21}\mathbf{W}_{11}^{-2}\mathbf{W}_{12} + 1) \\ &= \tilde{t}_p^2\omega^{pp}\big(\mathbf{W}_{21}\mathbf{W}_{11}^{-1/2}\mathbf{W}_{11}^{-1}\mathbf{W}_{11}^{-1/2}\mathbf{W}_{12} + 1\big). \end{aligned}$$

It is well known that ω^{pp} is independent of $\mathbf{W}_{21}\mathbf{W}_{11}^{-1/2}\mathbf{W}_{11}^{-1} \times \mathbf{W}_{11}^{-1/2}\mathbf{W}_{12}$ and $(m-2)/\omega^{pp} \sim \chi^2_{m-p-1}$ (Anderson 1984). Thus

$$\begin{aligned} \tilde{t}_p^2\omega^{pp} &= \frac{m-2}{m-p-1}\tilde{t}_p^2 \\ &\qquad /\big\{[(m-2)/\omega^{pp}]/(m-p-1)\big\} \\ &\sim \frac{m-2}{m-p-1}F_1. \end{aligned}$$

Note that $\mathbf{W}_{21}\mathbf{W}_{11}^{-1/2} \sim \mathrm{N}_{p-1}(0, \mathbf{I})$, $(m-2)\mathbf{W}_{11} \sim \mathrm{Wishart}_{p-1}(m-2, \mathbf{I})$, and they are independent (Anderson 1984). According to the relationship between Hotelling's T^2 distribution and the F distribution, it can be immediately obtained that $\mathbf{W}_{21}\mathbf{W}_{11}^{-2}\mathbf{W}_{12} \sim \frac{p-1}{m-p}F_2$, which completes the proof.

Proof of Proposition 4

Csorgo and Horvath (1997) presented a very general proof for the classical likelihood ratio approach. To apply their proofs, we need to make some minor modifications to our problem so that it is of the same form as that of Csorgo and Horvath (1997, sec. 1.1). Note that for $k = 1, \ldots, p$, the mean of $\mathbf{y}_j$ can be expressed as $\boldsymbol{\mu}_0 + \delta_j\mathbf{d}_k, j = 1, \ldots, m$. Then our hypothesis is equivalent to

$$H_0: \delta_1 = \delta_2 = \cdots = \delta_m$$

against the alternative

$$H_A: \text{there is an integer } \tau \text{ so that } \delta_1 = \cdots = \delta_\tau \neq \delta_{\tau+1} = \cdots = \delta_m.$$

This is exactly a one-dimensional parameter change point hypothesis. $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ are the nuisance parameters. It can be easily seen that the statistics V_k are similar to those seen when using the exact parametric likelihood procedure, as illustrated by Csorgo and Horvath (1997), except that the estimate of nuisance parameter $\boldsymbol{\Sigma}$ in V_k is obtained by the two-sample pooled sample variance matrix, $\mathbf{W}_j$, but the likelihood procedure uses the standard ML estimator by solving the likelihood estimating equations. However, as we know, when the process is statistically IC (under null hypothesis H_0), $\mathbf{W}_j$ is $m^{1/2}$-consistent, [i.e., $\|\mathbf{W}_j - \boldsymbol{\Sigma}\| = O_p(m^{-1/2})$], as is the estimate obtained from the estimating equations. Thus we need only replace the estimate of $\boldsymbol{\Sigma}$ with $\mathbf{W}_j$ in the proof of theorem 1.3.1 of Csorgo and Horvath (1997), which can similarly complete the remainder of the proof.

[Received April 2007. Revised January 2008.]

REFERENCES

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis* (2nd ed.), New York: Wiley.

Basseville, M., and Nikiforov, I. V. (1993), *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, NJ: Prentice-Hall.

Bhattacharya, P. K. (1987), "Maximum Likelihood Estimation of a Change Point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis*, 23, 183–208.

Camelio, J., Hu, S. J., and Ceglarek, D. (2003), "Modeling Variation Propagation of Multi-Station Assembly Systems With Compliant Parts," *Journal of Mechanical Design*, 125, 673–681.

Chen, J., and Gupta, A. K. (2000), *Parametric Statistical Change Point Analysis*, Berlin: Birkhauser.

Csorgo, M., and Horvath, L. (1997), *Limit Theorems in Change-Point Analysis*, New York: Wiley.

Ding, Y., Ceglarek, D., and Shi, J. (2002a), "Fault Diagnosis of Multistage Manufacturing Processes by Using a State-Space Approach," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 124, 313–322.

Ding, Y., Jin, J., Ceglarek, D., and Shi, J. (2005), "Process-Oriented Tolerancing for Multi-Station Assembly Systems," *IIE Transactions*, 37, 493–508.

Ding, Y., Shi, J., and Ceglarek, D. (2002b), "Diagnosability Analysis of Multistage Manufacturing Processes," *ASME Transactions, Journal of Dynamic Systems, Measurement and Control*, 124, 1–33.

Djurdjanovic, D., and Ni, J. (2001), "Linear State-Space Modeling of Dimensional Machining Errors," *Transactions of NAMRI/SME*, XXIX, 541–548.

Durbin, J., and Koopman, S. J. (2001), *Time Series Analysis by State-Space Methods*, New York: Oxford University Press.

Gombay, E., and Horvath, L. (1996), "Approximations for the Time and Change and the Power Function in Change Point Models," *Journal of Statistical Planning and Inference*, 52, 43–66.

Hawkins, D. M. (1977), "Testing a Sequence of Observations for a Shift in Location," *Journal of the American Statistical Association*, 72, 180–186.

Huang, Q., Zhou, S., and Shi, J. (2002), "Diagnosis of Multi-Operational Machining Processes Through Variation Propagation Analysis," *Robotics and CIM Journal*, 18, 233–239.

James, B., James, K. L., and Siegmund, D. (1987), "Tests for a Change Point," *Biometrika*, 74, 71–84.

——— (1992), "Asymptotic Approximations for Likelihood Ratio Tests and Confidence Regions for a Change Point in the Mean of a Multivariate Normal Distribution," *Statistica Sinica*, 2, 69–90.

Jin, J., and Shi, J. (1999), "State-Space Modeling of Sheet Metal Assembly for Dimensional Control," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 121, 756–762.

Kim, T. S., and May, G. S. (1999), "Sequential Modeling of via Formation in Photosensitive Dielectric Materials for MCM–D Applications," *IEEE Transactions on Semiconductor Manufacturing*, 12, 345–352.

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: Wiley.

——— (1991). **???**.

Rao, C. R., and Kleffe, J. (1988), *Estimation of Variance Components and Applications*, Amsterdam: North-Holland.

Rao, P. S. R. S. (1997), *Variance Components Estimation: Mixed Models, Methodologies and Applications*, London: Chapman & Hall.

Roy, S. N. (1953), "On a Heuristic Method of Test Construction and Its Use in Multivariate Analysis," *The Annals of Mathematical Statistics*, 24, 220–238.

Samuel-Cahn, E. (1996), "Is the Simes Improved Bonferroni Procedure Conservative?" *Biometrika*, 83, 928–933.

Sarkar, S. K. (1998), "Some Probability Inequalities for Ordered MTP_2 Random Variables: A Proof of Simes's Conjecture," *The Annals of Statistics*, 26, 494–504.

Sarkar, S. K., and Chang, C. K. (1997), "The Simes Method for Multiple Hypotheses Testing With Positively Dependent Test Statistics," *Journal of the American Statistical Association*, 92, 1601–1608.

Shi, J. (2006), *Stream of Variation Modeling and Analysis for Multistage Manufacturing*, **???**: Taylor and Francis.

Simes, R. J. (1986), "An Improved Bonferroni Procedure for Multiple Tests of Significance," *Biometrika*, 73, 751–754.

Srivastava, M. S., and Worsley, K. J. (1986), "Likelihood Ratio Tests for a Change in the Multivariate Normal Mean," *Journal of the American Statistical Association*, 81, 199–204.

Sullivan, J. H., and Woodall, W. H. (2000), "Change Point Detection of Mean Vector or Covariance Matrix Shifts Using Multivariate Individual Observations," *IIE Transactions*, 32, 537–549.

Vostrikova, L. J. (1981), "Detecting Disorder in Multidimensional Random Processes," *Soviet Mathematics Doklady*, 24, 55–59.

Xiang, L., and Tsung, F. (2006), "Statistical Monitoring of Multistage Processes Based on Engineering Models," *IIE Transactions*, **???**.

Yao, Y. (1988), "Estimating the Number of Change-Points via the Schwarz Criterion," *Statistics and Probability Letters*, 6, 181–189.

Zantek, P. F., Gordon, P. W., and Robert, D. P. (2006), "A Self-Starting Procedure for Monitoring Process Quality in Multistage Manufacturing Systems," *IIE Transactions*, 38, 293–308.

Zhou, S., Chen, Y., and Shi, J. (2004), "Statistical Estimation and Testing for Variation Root-Cause Identification of Multistage Manufacturing Processes," *IEEE Transactions on Automation Science and Engineering*, 1, 73–83.

Zhou, S., Ding, Y., Chen, Y., and Shi, J. (2003a), "Diagnosability Study of Multistage Manufacturing Processes Based on Linear Mixed-Effects Models," *Technometrics*, 45, 312–325.

Zhou, S., Huang, Q., and Shi, J. (2003b), "State-Space Modeling of Dimensional Variation Propagation in Multistage Machining Process Using Differential Motion Vector," *IEEE Transactions on Robotics and Automation*, 19, 296–309.